



UNIVERSITÀ DI PISA

CORSO DI DOTTORATO  
IN MICROBIOLOGIA E GENETICA

Anno di inizio del corso di dottorato: 2006

**“Polymorphic micro-RNA targets and risk of  
colorectal cancer”**

RELATORE:

Dr. Stefano Landi

CANDIDATA:

Dr. Debora Landi

# INDEX

<b>ABSTRACT</b>	<b>4</b>
<b>1. INTRODUCTION</b>	<b>6</b>
<b>1.1. MicroRNA</b>	<b>6</b>
<i>1.1.1. What is a microRNA?</i>	6
<i>1.1.2. MicroRNA discovery</i>	6
<i>1.1.3. Genomic localization</i>	8
<i>1.1.4. MicroRNA biogenesis</i>	9
<i>1.1.5. MicroRNA function</i>	15
<i>1.1.6. MicroRNAs and siRNAs</i>	18
<i>1.1.7. Polymorphisms and microRNA</i>	19
<i>1.1.8. Targets prediction and identification of new miRNAs</i>	21
<i>1.1.9. MicroRNAs and cancer</i>	22
<b>1.2. Sporadic Colorectal Cancer</b>	<b>26</b>
<i>1.2.1. Colorectal cancer epidemiology</i>	26
<i>1.2.2. Risk factors</i>	28
<i>Obesity, physical activity and inflammation</i>	28
<i>Diet</i>	30
<b>1.3. Aim</b>	<b>34</b>
<b>2. MATERIALS AND METHODS</b>	<b>36</b>
<b>2.1. Selection of genes</b>	<b>36</b>
<b>2.2. Targets prediction</b>	<b>36</b>
<b>2.3. Selection of polymorphisms</b>	<b>38</b>
<b>2.4. Genotyping</b>	<b>40</b>
<i>2.4.1. Study populations</i>	40
<i>Czech Republic</i>	40
<i>Spain</i>	42
<i>Italy</i>	43

German	44
Interviews	45
2.4.2. Laboratory Techniques	45
TaqMan allelic discrimination	45
Oligonucleotide allele specific PCR (ASO-PCR)	48
Fragment Analysis	50
2.4.3. Statistical Analysis	52
<b>3. RESULTS</b>	<b>53</b>
<b>4. DISCUSSION</b>	<b>56</b>
<b>5. CONCLUSIONS</b>	<b>61</b>
<b>6. FUTURE PROSPECTIVES</b>	<b>62</b>
<b>APPENDIX</b>	<b>63</b>
Table1	63
Table2	64
Table3	65
Table4	65
Table5	66
Table6	67
Table7a	68
Table7b	69
Table8	70
Table9	71
Table10	72
Table11	73
<b>BIBLIOGRAPHY</b>	<b>74</b>

# ABSTRACT

Recent evidence indicate that small non-coding RNA molecules, called micro-RNAs (miRNA), can bind to the 3'UTRs of mRNAs and interfere with their translation, thereby regulating cell growth, differentiation, apoptosis, and tumorigenesis. Genetic polymorphisms can reside on miRNA binding sites. Thus, it is conceivable that the miRNA regulation may be affected by polymorphisms on the 3' UTRs. Since gene de-regulation is one of the key mechanisms by which cells can progress to cancer, we hypothesize that common polymorphisms within miRNA target binding sites could play a role in the individual risk of cancer.

In the present study, we selected the 3'UTR regions of 129 genes candidate for colorectal cancer (CRC) and we identified putative miRNA binding sites by specialized algorithms (PicTar, DianaMicroT, miRBase, miRanda, TargetScan, and microInspector). We evaluated the SNPs for their ability to affect the binding of the miRNA with its target, by assessing the variation of Gibbs free energy between the two alleles of each SNP. We found 15 common polymorphisms. We added to this list 8 SNPs in miRNA sequences. All the polymorphisms were further investigated by a case-control association studies. The study was carried out on a series of cases and controls from Czech Republic, a population with the highest worldwide incidence of CRC. We found statistically significant associations between risk of CRC and variant alleles of *CD86* (OR=2.74 95%CI=1.24-6.04, for the variant homozygotes) and *INSR* genes (OR=1.94; 95%CI=1.03-3.66, for the variant homozygotes).

Then, these two polymorphisms were genotyped in three different populations: Spanish, Italian, and German. The statistical analyses for all the samples (Czech, Spanish, Italian, and German) confirmed the association between risk of CRC and the polymorphisms in *CD86* and *INSR*. These results are the first reporting positive association between miRNA-binding SNPs sequences and cancer risk.

# **1. INTRODUCTION**

## **1.1. MicroRNA**

### **1.1.1. What is a microRNA?**

MicroRNAs (miRNAs) are endogenous non-coding RNAs of about 22 nucleotides, regulating genes in animals and plants by pairing to the 3'UTR regions of messenger RNAs (Zeng et al., 2002) of target genes and specifying mRNA cleavage or repression of protein synthesis. There is increasing evidence that miRNAs have an important regulatory role in a broad range of biological processes, including developmental timing, cellular differentiation, proliferation, apoptosis, cancer development, insulin secretion, and cholesterol biosynthesis (Bartel, 2004b). After the initial genetic discovery in nematodes (Lee et al., 1993a), it was soon confirmed that all investigated higher eukaryotes, including plants and mammals, contain miRNA genes (Ambros et al., 2003b). MiRNA genes represent only a small part of the genome (3%), but they regulate approximately from 20% to 30% of all human genes and there is an average of 200 predicted targets per miRNA (Carthew, 2006).

### **1.1.2. MicroRNAs discovery**

This discovery has been an exciting breakthrough in Biological Sciences of the past decade, culminating in Nobel Prize in Physiology or Medicine awarded to Andrew Fire and Craig Mello. Building on previous work mainly in plants (Lindbo et al., 1993), Fire discovered that exogenous double-stranded RNA can be used to specifically interfere with gene function. This phenomenon

was called RNA interference (RNAi) (Fire et al., 1998). They also speculated that organisms might use double-stranded RNA naturally as a way of silencing genes. It was then shown that RNA interference was mediated by 22 nucleotide single-stranded RNAs termed small interfering RNAs (siRNAs) derived from the longer double-stranded RNA precursors (Zamore et al., 2000). Over the following years, many new small functional RNAs have been found. The first miRNA was discovered in 1993 by Victor Ambros and colleagues Rosalind Lee and Rhonda Feinbaum. A genetic screen in the groundworm *Caenorhabditis elegans*, a millimeter-long animal used as a model organism in biological research, identified genes involved in developmental timing (Lee et al., 1993b). Surprisingly, one of the genes, termed *lin-4*, did not encode a protein but instead a novel 22-nucleotide small RNA. Seven years later, Reinhart et al. discovered a second 22-nucleotide small RNA of this type, *let-7*, a gene also involved in *C. elegans* developmental timing.

The homologs of the *let-7* gene were identified in other animals including humans (Pasquinelli et al., 2000). The conservation of *let-7* across species suggested an important and fundamental biological role for this small RNA.

The mechanism of RNA interference (RNAi) was discovered at that time, and it became clear that miRNA and RNAi pathways were intricately linked and shared common components. Within the following year, more than 100 additional small regulatory RNAs similar to *lin-4* and *let-7* were identified in worms, the fruit fly *Drosophila*, and in humans (Lee and Ambros, 2001). Subsequently, many more short regulatory RNAs were identified in almost all multicellular organisms, including flowering plants, worms, flies, fish, frogs,

mammals (Lim et al., 2003), and in single cellular algae and DNA viruses (Sullivan and Ganem, 2005).

To date, more than 500 human miRNAs have been experimentally identified. Computational predictions of miRNA targets suggest that up to 30% of human protein coding genes may be regulated by miRNAs (Rajewsky, 2006b). This makes miRNAs one of the most abundant classes of regulatory genes in humans. MicroRNAs are now perceived as a key layer of post-transcriptional control within the networks of gene regulation.

### **1.1.3. Genomic localization**

Micro-RNAs are initially transcribed as precursor molecules in the nucleus, where many are organized in genomic cluster, while others exist as individual genes. Besides miRNAs that are non linked to any other transcription unit, it has been reported that miRNA genes are located in exons or introns of non-coding genes (Rodriguez et al., 2004), as well as protein-coding genes (Smalheiser, 2003). It is not yet clear if these miRNAs are functionally related to their host genes, since the splicing machinery may not be able to release an intron that is accessible to further miRNA processing. On the other hand, splicing and miRNA processing might be coupled and miRNAs and their host mRNAs could be processed simultaneously. However, expression studies on miRNAs and host mRNAs revealed that miRNAs are frequently co-expressed with their host genes (Baskerville and Bartel, 2005). A sizable minority (e.g., about a quarter of the human miRNAs genes) are in the introns of pre-mRNAs. These are preferentially in the same orientation as the predicted mRNAs, suggesting that most of these miRNAs are not transcribed from their own



promoters but are instead processed from introns. This arrangement provides a convenient mechanism for the coordinated expression of a miRNA and a protein (Lagos-Quintana et al., 2003).

Some mammalian miRNA genes are located in repetitive genomic regions as well, and it has been suggested that repetitive elements such as transposons may be the driving force that creates new miRNAs during mammalian evolution (Smalheiser and Torvik, 2005).

MicroRNAs can be grouped into families on the basis of sequence homology, which is found primarily at the 5' end of the mature miRNAs, but whether members of the same miRNA family control similar biological events remains to be seen. Many miRNAs are evolutionarily conserved from worms to humans, which implies that these miRNAs direct essential processes both during development and in the adult body (Esquela-Kerscher and Slack, 2006).

#### **1.1.4. MicroRNA biogenesis**

The maturation of microRNA is a multi-step process that begins in the nucleus and ends in the cytoplasm (Novina and Sharp, 2004) (figure 1). Analyses have shown that human primary miRNA transcripts (pri-miRNAs) contain cap structures as well as poly (A) tails which are the unique properties of class II gene transcripts. These data indicate that pol II is the main RNA polymerase for miRNA gene transcription (Lee et al., 2004). Moreover, in 2006, it was reported that ~ 50 human miRNAs are transcribed by RNA polymerase III. In fact, in a genomic analysis, the miRNAs in the human chromosome 19 miRNA clusters were found to be dispersed among Alu repeats. Alu sequence is the most abundant transposable element in the human genome. It is derived

from the 7SL RNA gene, which encode the RNA component of the signal recognition particle that functions in protein synthesis. The Alu sequence contains the 7SL promoter, an RNA polymerase III promoter (Borchert et al., 2006).

Animal miRNAs are initially transcribed as part of a long RNA precursor, defined as “primary miRNA” (pri-miRNA). Within the pri-miRNA, the ~22-nt mature miRNA forms part of one arm of a ~80-nt imperfect stem-loop sequence (Cullen, 2004). The first step is the nuclear cleavage of the pri-miRNA, which liberates a ~60-70 nt stem loop intermediate, known as the miRNA precursor, or the pre-miRNA (Zeng, 2006). This processing is performed by Drosha, an RNase III-type endonuclease that contains two RNase III domains and a dsRNA binding domain in the C-terminal half, a proline-rich domain and an arginine-rich (R-rich) domain in the N-terminal half of the protein. It cleaves both strands of stem at sites near the base of the primary stem loop, leaving a 5' phosphate and a 2-nt 3' overhang (Basyuk et al., 2003). It has been demonstrated that for a pri-miRNA to be efficiently processed by Drosha the targeted hairpin must consist of a large terminal loop of  $\geq 10$  nt and a stem region somewhat longer than the one present in the final pre-miRNA (Zeng and Cullen, 2005). The class 3 RNase III Drosha forms a complex (microprocessor) with a double-stranded RNA-binding protein, termed DGCR8 in humans and Pasha in flies and worms (Han et al., 2004). Cleavage of a pri-miRNA by microprocessor begins with DGCR8 recognizing the ssRNA-dsRNA junction typical of a pri-miRNA. Then, Drosha is brought close to its substrate through interaction with DGCR8 and cleaves the stem of a pri-miRNA ~11 nt away from the two single stranded segments (Han et al., 2006). Although microprocessor is already sufficient for

conversion of a pri-miRNA into a pre-miRNAs *in vitro*, cleavage of pri-miRNA *in vivo* does not depend on Drosha and DGCR8 only, but also on other accessory proteins, such as the RNA binding protein hnRNP A1 and the p68 and p72 RNA helicases. Gene-targeting experiments have demonstrated the importance of p68 and p72 for the biogenesis of a subset of mouse miRNAs. In p68 or p72 knock-out embryos (knock-out of p68 in mice causes embryonic lethality while knock-out of p72 causes neonatal death), the expression of a subset of miRNAs is severely compromised. In another recent study, hnRNP A1 has been shown to be specifically required for the processing of pri-miR-18a in a context-dependent manner (Fukuda et al., 2007; Guil and Caceres, 2007). In addition to the classical biogenesis pathway that depends on microprocessor, a subclass of pre-miRNAs, pre-miRNA/introns (mirtrons), have recently been shown to depend on the RNA splicing machinery for their biogenesis in *Drosophila*, *Caenorhabditis elegans* and mammals (Berezikov et al., 2007). Mirtrons are derived from certain debranched introns that fold into hairpin structures with 5' monophosphates and 3' 2-nt hydroxyl overhangs, which mimic the structural hallmarks of pre-miRNAs and enter the miRNA processing pathway (Okamura et al., 2007; Ruby et al., 2007). The discovery of mirtrons suggests that any RNA, with a size comparable to a pre-miRNA and all the structural features of a pre-miRNA, can be utilized by the miRNA processing machinery, and potentially give rise to a functional miRNA.

After initial nuclear processing, the pre-miRNA is exported to the cytoplasm by the export receptor Exportin-5 (Exp-5). This step is a Ran-dependent and requires the hydrolysis of GTP (Yi et al., 2003). Although Exportin-5 lacks a structurally known dsRNA-binding domain, it has been

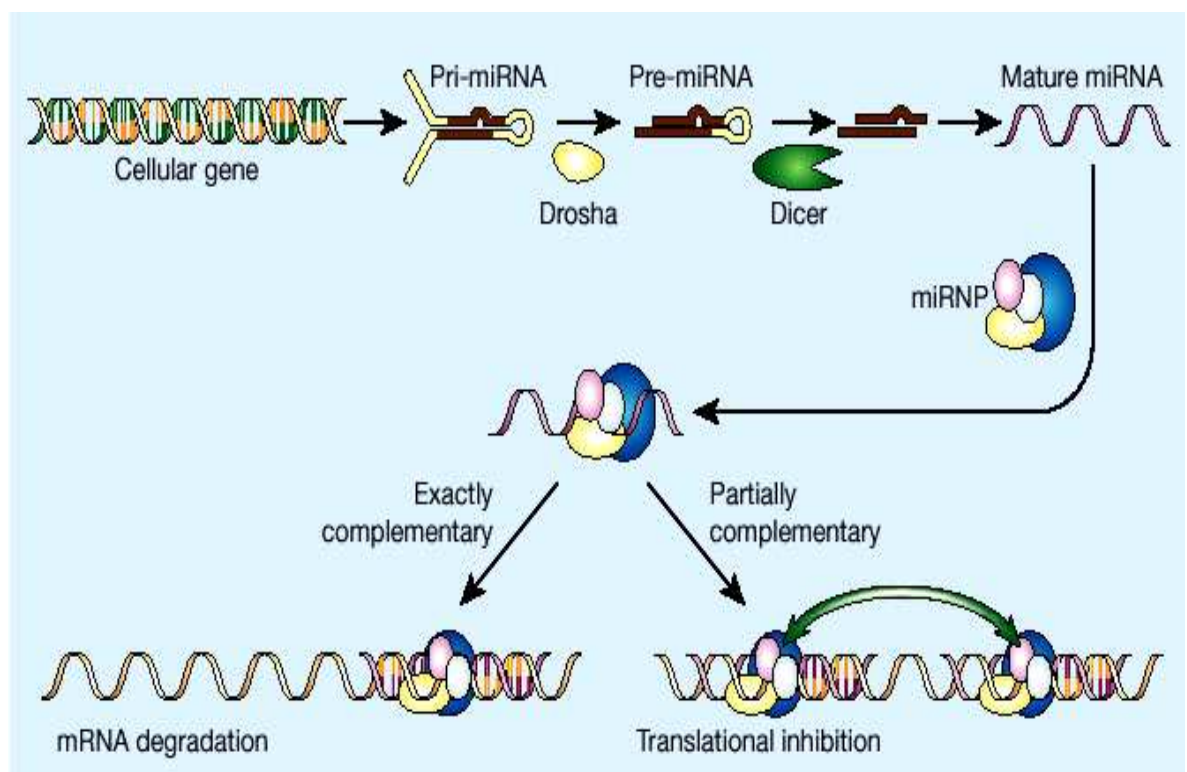
suggested that it interacts directly with the pre-miRNA. The interaction of Exportin-5 with the miRNA precursor requires a 3' overhang and the stem of the precursor for efficient export (Zeng and Cullen, 2004). Exp-5 is also important for stabilizing pre-miRNAs in the nucleus. When Exp-5 is knocked down by siRNAs, the levels of pre-miRNAs are reduced not only in the cytoplasm, but also in the nucleus, suggesting that binding of pre-miRNAs to Exp-5 protects them from degradation (Yi et al., 2003).

In the cytoplasm, another RNase III endonuclease termed Dicer is responsible for dicing pre-miRNAs into short RNA duplexes termed miRNA duplexes (Bernstein et al., 2001). The RNA strand of the miRNA duplex that is complementary to the mature miRNA is depicted with a star symbol (miRNA\*). In addition to two RNase III signature domains, mammalian Dicer has a N-terminal ATPase/helicase domain, a DUF 283 domain, a PAZ domain and a C-terminal dsRNA binding domain (dsRBD) (Provost et al., 2002). Biochemical experiments have revealed that both PAZ domain and dsRBD are essential for the interaction of Dicer with pre-miRNAs and long dsRNAs. PAZ domain functions in recognizing the 2-nt 3' overhang signature generated by Drosha, while dsRBD is critical for binding long dsRNAs. After capturing pre-miRNAs or long dsRNAs with its PAZ domain and dsRBD, Dicer dimerizes its two RNase III domains intra-molecularly to form a single processing center, and cuts the stem of pre-miRNAs or long dsRNAs ~22 nt away from their termini at positions separated by 2 nts, which generates 3' 2-nt termini (Zhang et al., 2004). In the case of long dsRNA, subsequent to the first cleavage, Dicer can use its PAZ domain to place it again at the termini of the dsRNA and dicer processively from the termini (Zhang et al., 2002). Although the functions of the C-terminal

domains of Dicer are clear now, the roles of the N-terminal ATPase/helicase domain and DUF 283 domain of Dicer still remain elusive.

After Dicer processing, the miRNA duplex is unwound and the mature miRNA binds to an Argonaute (Ago) protein in a process that is referred to as miRNA loading or assembly, while the miRNA\* is degraded. The Argonaute family is a diverse family of proteins, each containing characteristic domains termed PAZ and PIWI. The Argonaute family can be phylogenetically divided into the Ago and Piwi protein families based on similarities to *Arabidopsis* AGO1 and *Drosophila* Piwi proteins, respectively (Carmell et al., 2002). miRNAs bind Ago proteins whereas Piwi proteins bind a newly discovered class of small RNAs known as piwi-interacting RNAs (piRNAs), which are almost exclusively expressed in the germline (Aravin et al., 2007; Kim, 2006). Humans and mice contain four Ago proteins (Ago1–4). Structural and biochemical analyses have shown that the ~130- amino-acid PAZ domain contains an oligonucleotide-binding fold that allows the protein to bind the single-stranded 2-nt 3' terminal overhangs characteristic of small RNAs processed by Dicer (Lingel et al., 2004). The miRNA/Ago ribonucleoprotein that is formed represents the core component of the effector complexes that mediate miRNA function and is known as miRNP (Mourelatos et al., 2002). A primary determinant of which the two strands of a miRNA duplex or a siRNA duplex will be loaded on Ago proteins is the inherent thermodynamic asymmetry of the miRNA or siRNA duplex. The RNA strand whose 5' end is less stably bound to the opposite strand will be loaded to Ago proteins and forms the mature miRNA or siRNA (Tomari and Zamore, 2005). Humans and other mammals contain a single Dicer gene and miRNP, and RISC assembly has many similarities but also

important differences to RISC assembly in flies. In humans, miRNP assembly is accomplished by a protein complex termed the miRNA RISC Loading Complex (miRLC). The miRLC is a multiprotein complex whose core components are Ago and Dicer proteins (Gregory et al., 2005). The miRLC is devoid of miRNAs and processes miRNAs from pre-miRNAs, and loads mature miRNAs to Ago proteins. The miRLC is then disassembled and the core miRNP (miRNA-Ago ribonucleoprotein) is generated (Maniataki and Mourelatos, 2005). However, the details of miRNP assembly in humans are unknown.



**Figure 1: microRNA biogenesis (Novina and Sharp, 2004)**

### 1.1.5. MicroRNA function

MiRNAs base-pair with miRNA recognition elements (MREs) found in their mRNA targets (typically in the 3' untranslated region-3'UTR) and deposit their bound Ago proteins onto mRNA targets. The result is translational repression of the targeted mRNA, often followed by mRNA destabilization or endonucleolytic cleavage of the targeted mRNA. The exact molecular function is dependent upon how extensive the complementarity of the miRNA or siRNA is with its mRNA target and which Ago protein is deposited on the mRNA target. If an miRNA or siRNA bound to Ago2 pairs with extensive complementarity with a cognate mRNA target, then the mRNA is cleaved at a position across from nucleotides 10 and 11 of the miRNA (or siRNA), while the miRNA remains intact (Figure 2) (Liu et al., 2004). This cleavage event produces 5'-phosphate and 3'-hydroxyl terminal products, characteristic of other RNase H-like enzymes (Martinez and Tuschl, 2004). The target mRNA is subsequently degraded via routine cellular pathways (Figure 2). Target mRNA cleavage by miRNAs is the major mechanism of regulation by plant miRNAs (Dugas and Bartel, 2004).

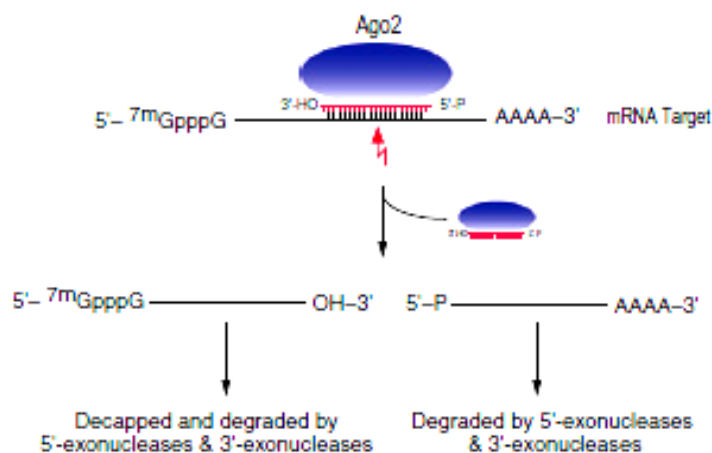
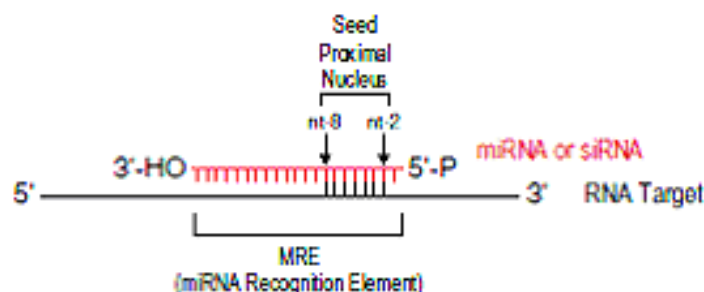


Figure 2: target RNA cleavage by Ago2-containing miRNP (Liu et al, 2007).

In animals, however, there are very few examples of miRNAs that regulate their mRNA targets by cleavage (Yekta et al., 2004); rather, the predominant silencing mode of animal miRNAs is to repress the translation of their mRNA targets and/or to destabilize them without endonucleolytic cleavage (Filipowicz et al., 2005). Experimental and bioinformatics approaches have shown that the most important determinant of target RNA recognition by a miRNA is perfect or near-perfect complementarity between the proximal (5') region of the miRNA and the mRNA, also known as the “seed” region or the “nucleus” (Figure 3) (Lai, 2002). Base-pairing between the 3' portion of the miRNA and the mRNA target is not always essential for repression, but strong base-pairing within this region can partially compensate for weaker seed matches or enhance repression (Brennecke et al., 2005). Additionally, multiple MREs for the same, or different, miRNAs within the same 3'UTR can function cooperatively to enhance repression (Krek et al., 2005). Spacing of the seed sites within the 3'UTR may play a significant role in the cooperative action of miRNAs. Finally, sequences adjacent to MREs (Vella et al., 2004) and the secondary structure of the 3'UTR of the target mRNA affect binding of miRNAs.

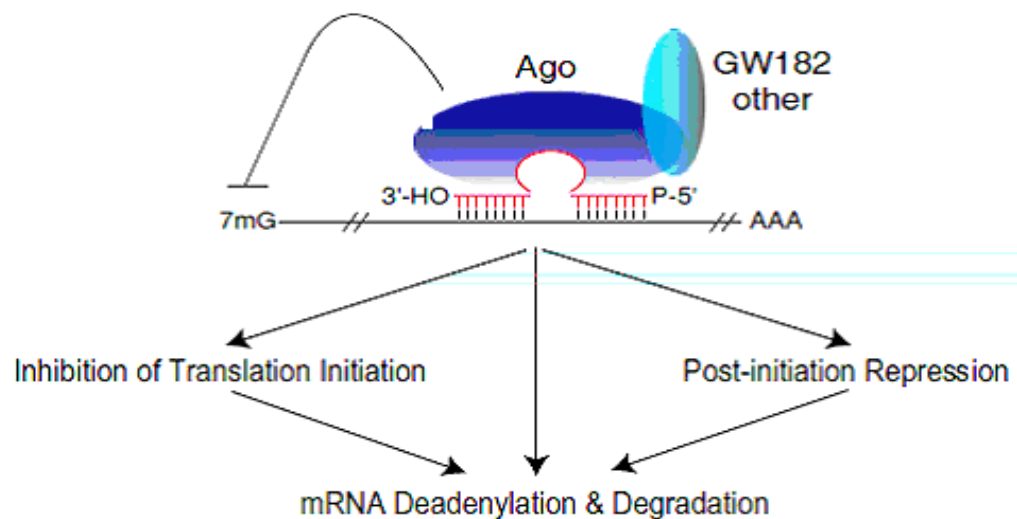


**Figure 3: Principles of miRNA binding to target RNA (Liu et al, 2007)**



Initial studies suggested that miRNA-mediated translational repression occurred at a step following the initiation of translation. This was based on the observation that, in *C. elegans*, *lin-14* mRNA, a target of the *lin-4* miRNA, could be found associated with mRNAs in polysomes, according to sucrose gradient sedimentation analysis (Seggerson et al., 2002). MiRNAs also associate with polysomes in both *C. elegans* and mammalian cells (Nelson et al., 2004). More recent studies have lent additional support to the model of miRNA mediated repression occurring at a post-initiation step (Petersen et al., 2006) (Figure 4).

MiRNA mediated translational repression results in decreased levels of the targeted protein. Thus, the model of miRNA function occurring at a post-initiation step would require the release of and subsequent destruction of the resulting polypeptide (Nottrott et al., 2006).



**Figure 4: Mechanism of repression of targeted mRNA by miRNPs (Liu et al, 2007)**

A second model proposes that miRNA-mediated translational repression occurs at the initiation step (Figure 7B). Using human cells and reporter constructs targeted by either endogenous or exogenous miRNAs, two groups initially discovered that the m7G cap of an mRNA is required for efficient translational repression (Humphreys et al., 2005). MicroRNA-repressed messages sedimented in light fractions in polysome analyses, indicating that their translation was inhibited at the level of initiation (Pillai et al., 2005).

#### **1.1.6. MicroRNAs and siRNAs**

Important distinctions between miRNA and siRNA can be made, particularly in regard to their origin, evolutionary conservation, and the types of genes that they silence.

In fact, microRNAs derive from genomic loci distinct from other recognized genes and are processed from transcripts that can form local RNA hairpin structures, whereas siRNAs often derive from mRNAs, transposons, viruses, or heterochromatic DNA and are processed from long bimolecular RNA duplexes or extended hairpins. A single miRNA: miRNA\* duplex is generated from each miRNA hairpin precursor molecule, on the contrary a multitude of siRNA duplexes are generated from each siRNA precursor molecule, leading to many different siRNAs accumulating from both strands of this extended dsRNA. Moreover, miRNA sequences are nearly always conserved in related organisms, whereas endogenous siRNA sequences are rarely conserved (Ambros et al., 2003a).

Endogenous siRNAs typically specify “auto-silencing”, in that they specify the silencing of the same locus (or very similar loci) from which they originate,

whereas miRNAs specify “hetero-silencing”, in that they are produced from genes that specify the silencing of very different genes. Natural examples of auto-silencing include the silencing of virus, transposons, and the heterochromatic outer repeats of centromeres (Aravin et al., 2001).

To the extent that the siRNAs come from the same loci that they target, a mutational event that changes the sequence of the siRNA would also change the sequence of its regulatory target, and siRNA regulation would be preserved. In contrast, a mutation in a miRNA would rarely be accompanied by simultaneous compensatory changes at the loci of its targets, and thus selection pressure would preserve the miRNA sequence (Bartel, 2004a).

#### **1.1.7. Polymorphisms and microRNA**

Single nucleotide polymorphisms (SNPs) are the most abundant form of DNA variation in the human genome and contribute to human phenotypic differences. Polymorphisms in miRNA genes could potentially alter various biological processes by influencing the processing and/or target selection of miRNAs. Duan and colleagues have identified 323 SNPs in 227 known miRNAs examined: 12 of these are located within the miRNA precursors. Interestingly, a polymorphism (G/U) was identified at the eighth nucleotide of mature miR-125a. This miR-125a SNP blocks the processing of pri-miRNA to pre-miRNA significantly in addition to reducing miRNA-mediated translational suppression. These data suggest that SNPs that reside within the miRNA genes could indeed regulate miRNA biogenesis and alter target selection, thereby potentially having profound biological effects (Duan et al., 2007). In the human genome, miR-125a is located at chromosome 19q13.41, a region that is frequently deleted in

primary gliomas, especially oligodendrogliomas (Law et al., 2005). Identified targets of miR-125a include *Lin-28*, *Lin-41*, *ERBB2* and *ERBB3*. MiR-125a was found to be down-regulated in one breast cancer miRNA profiling study (Wu and Belasco, 2005).

In a study of Iwai e Naraba, 173 human pre-miRNA genome regions in 96 subjects were sequenced and 10 polymorphisms in 10 pre-miRNA hairpins were identified. Furthermore, a C to A polymorphisms in the mature miR-30c-2 sequence was also identified. They showed that mature microRNA production was highly dependent on the integrity of the precursor RNA stem, although the underlying specific sequence had little effect. They also showed that the specific sequence of the terminal loop only moderately affected microRNA production (Iwai and Naraba, 2005).

An association between microRNA-binding SNPs and the risk of common diseases was indicated in three different studies. In a landmark study, Abelson et al. (Abelson et al., 2005) showed that a SNP in the 3'UTR human *SLITRK1* gene strengthens an existing miR-189 target site, thereby amplifying the down-regulation of *SLITRK1*, which is implicated in Tourette syndrome. Another study demonstrated that a 3' UTR SNP in the sheep myostatin (*Gdf8*) gene creates a new illegitimate microRNA target site, which leads to the significant down-regulation of *Gdf8* and contributes to the development of muscular hypertrophy (Clop et al., 2006). In the third study, two groups almost simultaneously investigated the 3' UTR of the human *AGTR1* gene that contains the SNP rs5186 and showed that hsa-miR-155 specifically down-regulates the expression of only the 1166A, and not the 1166C, allele of rs5186 (Martin et al., 2007; Sethupathy et al., 2007). Both groups concluded that, by abrogating the

regulation by miR-155, the 1166C allele may be functionally associated with hypertension and cardiovascular disease.

#### **1.1.8. Targets prediction and identification of new miRNAs**

In the past few years many computational methods to identify the targets of the miRNAs were developed. These methods search regions conserved in the 3'UTR that are complementary to the micro-RNA. The identification of the messengers target is more difficult in the animals than in the plants, because in the animals there are less mRNA perfectly complementary to the miRNA (Rajewsky, 2006a). The analysis of the binding sites of the miRNAs, that they must be validated experimentally, is based on more criteria: a perfect complementarity between the region 3'UTR of mRNA target and the first 8 nucleotide of miRNA beginning from its extremity 5'UTR (inside of this region is admitted pairings G: U), the formation of a heteroduplex structurally and thermodynamically stable, and the evolutionary conservation of the sites target between the vertebrates (Brennecke et al., 2005). Several independent groups have formulated of the algorithms that serve to identify the targets of the miRNA (John et al., 2004). The binding of more miRNAs to the same messenger complicate the prediction of the targets (Krek et al., 2005).

In several organisms hundred of miRNAs were identified and successively same studies were carried out on the profiles of genic expression (Krutzfeldt et al., 2006). Some characteristics allow to define if the identified molecule is a true one micro-RNA: a miRNA mature must be express as a transcript of 22 nucleotides, derive from precursory with characteristic secondary structure (stem-loop structure), occupy the part of stem and be processed by Dicer.

Other criterion commonly used regards the conservation of the sequence of the micro-RNA and the stem-loop structure in various species. A “ideal” microRNA would have to satisfy all these criteria: only one of these criteria is not sufficient in order to assess that a candidate gene is a new miRNA (Ambros et al., 2003c).

### **1.1.9. MicroRNAs and cancer**

The first evidence for miRNA involvement in human cancer came from a study by Calin et al. (Calin et al., 2002). Examining a recurring deletion at chromosome 13q14 in the search for a tumor suppressor gene involved in chronic lymphocytic leukemia (CLL), this study found that the smallest minimal common region of deletion encodes two miRNAs, mir-15a and mir-16-1. Analysis of their expression in CLL samples and normal CD5+ lymphocytes revealed that down-regulation of miR-15a and miR-16-1, which shares a primary transcript, is consistently associated with the deletion at chromosome 13q14. This suggested a role of miR-15a and miR-16-1 as tumor suppressor genes. Subsequent investigations have confirmed the involvement of miRNAs in the pathogenesis of human cancer. The putative tumor suppressive role of miR-15a and miR-16-1 was supported by the discovery in two CLL patients of a germ-line point mutation that results in reduced levels of mature miR-15a and miR-16-1 (Calin et al., 2005), and the idea was further strengthened by the demonstration that miR-15a and miR-16-1 negatively regulate the anti-apoptotic oncogene *BCL2* at a post-transcriptional level and induce apoptosis in the leukemic cell line MEG-01 (Cimmino et al., 2005).

MicroRNA deregulation can also operate in the opposite direction in cancer. The miR-17-92 family is the most studied example. This family includes fourteen homologous miRNAs, which are encoded by three gene clusters on chromosomes 7, 13 and X (Tanzer and Stadler, 2004). The cluster on chromosome 13 is amplified in human B-cell lymphomas, which leads to increased expression of various miRNAs. Interestingly, enforced expression of the miR-17-92 cluster acts together with *MYC* to accelerate tumor development in a mouse B-cell lymphoma model (He et al., 2005); it thus acts as an oncogene. It has been reported that transcription of this cluster is also induced by *MYC* itself; oncogenic members of the miR-17-92 cluster may therefore act as *MYC* effectors. Bi-directional relationships between oncogene activation and miRNA deregulation thus exist. By inhibiting oncogenes or functioning as their effectors, miRNAs could themselves act as tumor suppressor genes or oncogenes (figure 5).

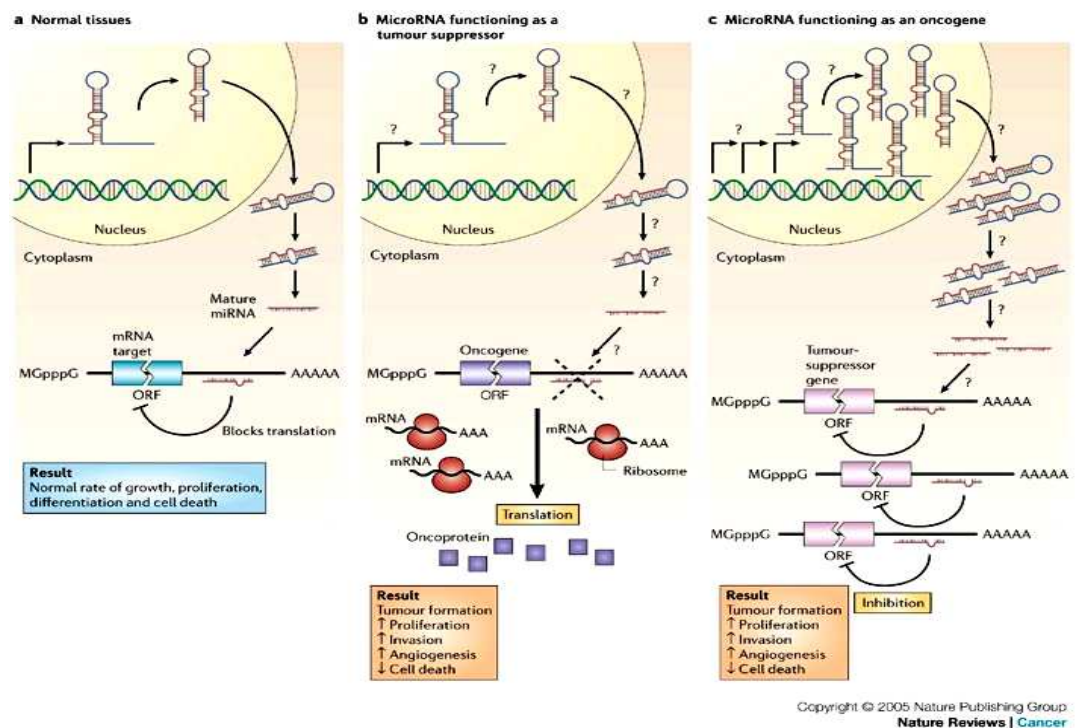


Figure 5: MicroRNA functioning as tumour suppressor (Liu et al, 2007)

Deregulation of miRNA expression levels emerges as the main mechanism that triggers their loss or gain of function in cancer cells. Genomic aberrations might alter miRNA expression, since miRNA up-regulation has been associated with genomic amplification (O'Donnell et al., 2005), and miRNA down-regulation has been associated with chromosomal deletions, point mutations and aberrant promoter methylation (Saito et al., 2006).

Several examples of miRNAs whose expression is deregulated in human cancer have been reported. Down-regulation of miR-143, miR-145 and members of the let-7 family has been described. The miR-143 and miR-145 genes, which reside in a genomic cluster similar to that encoding miR-15a/miR-16-1, are significantly down-regulated in colon cancer tissue compared with colonic mucosa (Michael et al., 2003). Let-7 family regulate the RAS oncogenes. The 3' UTR of the RAS mRNAs contains multiple binding sites for let-7 members, and forced expression of let-7 in human cancer cells reduces RAS protein levels (Johnson et al., 2005). Since let-7 is down-regulated in several human cancers, this mechanism could lead to the activation of the RAS pathway (Akao et al., 2006). Mir-125a and miR-125b are down-regulated in breast cancer (Iorio et al., 2005). These two miRNAs regulate the expression of the receptor tyrosine kinases ERBB2 and ERBB3 (Scott et al., 2007). Ectopic over-expression of miR-125a or miR-125b in SK-BR-3 cells induces impaired anchorage-dependent growth and reduced migration and invasion capacities, which is consistent with suppression of ErbB signalling.

An example of up-regulated miRNA is miR-155, which lies in the only phylogenetically conserved sequence of *BIC*, a non protein-coding gene that was discovered as site of pro-viral insertions in avian leucosis-virus-induced



lymphomas (Tam et al., 1997). This miRNA and its primary transcript *BIC* are over-expressed in Hodgkin lymphoma, in pediatric Burkitt lymphoma and in diffuse large B-cell lymphoma (Eis et al., 2005). Another example of up-regulated miRNA is miR- 21, a gene located at chromosome 17q23 in a chromosomal region frequently amplified in human cancer, which is up-regulated in human breast cancer and in glioblastoma (Chan et al., 2005).

A peculiar mechanism linking miRNA to cancer was revealed by the analysis of a masked chromosomal translocation t (8; 17) in a B-cell leukemia (Gauwerky et al., 1989), in which the mir-142 regulatory element is juxtaposed to the *MYC* proto-oncogene. Here, a miRNA regulatory element serves as a proto-oncogene activator. Although at present this mechanism appears to be unique, given the high levels and tissue-specific expression of several miRNAs it is a mechanism that should be further investigated in translocations identified in human cancer that are yet to be associated with culprit genes.

In most cases, deregulation consistently acts in one direction, either up-regulating or down-regulating the miRNAs, which suggests that these miRNAs are likely to play a crucial role in tumorigenesis. There are, however, some unusual situations: for example, members of the miR-181 family are up-regulated in some cancers, such as thyroid, pancreatic and prostate carcinomas but down-regulated in others, such as glioblastomas and pituitary adenomas (Pallante et al., 2006).

There are also examples of miRNAs deregulated in specific neoplasms: miR-122a, for example, a liver-specific miRNA, is down-regulated in hepatocellular carcinoma (Kutay et al., 2006); miR-204 and miR-211 are

specifically up-regulated in insulinomas (Roldo et al., 2006). Tissue-specific aberrant expression of miRNAs might reflect the differentiation status of the cell.

## **1.2. Sporadic Colorectal Cancer**

### **1.2.1. Colorectal cancer epidemiology**

Colorectal cancer (CRC) is the commonest cancer of the alimentary tract in the developed world, and it was the third most commonly diagnosed form of cancer in the UK in 2000. In the USA and Western Europe, Australia and New Zealand, the age-standardized rates of incidence of colorectal cancer among males are currently around 50 per 100 000, whereas in Bangladesh the rate is about 1 case per 100 000 (Parkin et al., 2005a). Migrants moving from regions of low incidence to developed countries tend to acquire the risk typical of the new host population within one generation. This provides strong circumstantial evidence to suggest that environmental factors, rather than genetic variations between populations, are of prime importance in the aetiology of the disease (Flood et al., 2000).

The sporadic form of colorectal carcinoma emerges principally via the adenoma-carcinoma sequence, in which malignant tumours develop from a small proportion of adenomatous polyps, probably over a period of decades. Populations with a high prevalence of sporadic adenomas also have a high incidence of colorectal cancer, and patients who harbour polyps have a greater risk of developing colorectal carcinoma than individuals with no previous history of polyps. The molecular basis for the emergence of the malignant phenotype is also relatively well established, at least in comparison with most other human cancers. The morphological changes of the adenoma-carcinoma sequence are

associated with a progressive acquisition of abnormalities of the genome, including somatic mutations affecting known proto-oncogenes or tumour-suppressor genes (Vogelstein et al., 1988).

For example, the so-called 'gatekeeper' mutation, a homozygous mutation of the APC gene, is thought to be amongst the earliest genetic changes in the adenoma-carcinoma sequence (Lamlum et al., 2000). Large adenomatous polyps and malignant tumours often have a mutation in the K-ras proto-oncogene, and mutations in the tumour suppressor gene p53 appear to be associated typically with the transition to a carcinoma (Gafa and Lanza, 1998). In addition to such somatic mutations, neoplastic progression in many tissues is also thought to be driven partially by epigenetic modifications to the expression of genes regulating proliferation, apoptosis and DNA repair. The most thoroughly studied of these effects is gene-silencing associated with the aberrant methylation of CpG islands in the promoterregions of genes. This mechanism does not involve any alteration to the DNA sequence, but the signal is nevertheless transmitted through mitosis (Esteller and Herman, 2002).

The major challenge facing those who seek to harness nutrition as a strategy for colorectal cancer prevention is the development of mechanistic hypotheses to account for interactions between diet and the disease process (Gunter and Leitzmann, 2006).

### **1.2.2. Risk factors**

- ***Obesity, physical activity and inflammation***

There is a consistent body of evidence from prospective studies to indicate that overweight and obesity are positively associated with risk of colon cancer and it have since been confirmed in prospective studies conducted in various parts of the world although the effects have generally been more clear-cut in men than in women (MacInnis et al., 2006).

Exercise was found to reduce the risk of cancer, but to have little influence on adenomas. High-energy intake increased the risk of cancer but not of adenomas, and high body-mass index (BMI) significantly increased the risk of large adenomas. Thus, the findings were consistent with an adverse effect of high body mass on the later stages of the adenoma-carcinoma sequence, though BMI was not shown to influence the risk of cancer (Samanic et al., 2006).

Abdominal visceral fat accounts for about 18% of total adipose tissue (Ross et al., 1992). Several recent studies have been designed to explore associations between the risk of colorectal cancer and high levels of abdominal fat. This adverse effect increased significantly with age and was stronger for men than for women (Moore et al., 2004). A larger waist size was associated with increased risk of colon cancer in both sexes, and this effect was independent of BMI. Again the risk was particularly strong amongst sedentary subjects. The overall conclusion was that around 2 h of moderate physical activity per day or 1 h of more vigorous activity was associated with a risk reduction for colon cancer of 20–25% (Pischon et al., 2006)

There is a strong and growing body of evidence to suggest that obesity and colon cancer are causally linked by mechanisms involving chronic, asymptomatic inflammatory activity in the colonic mucosa. Inflammatory bowel disease is a well-recognized cause of colorectal carcinogenesis (Munkholm, 2003).

Conversely, long-term consumption of aspirin and other non-steroidal anti-inflammatory drugs (NSAIDs) significantly reduces the risk of colorectal cancer, primarily through inhibition of the pro-inflammatory enzyme cyclooxygenase 2 (COX-2) (Tuynman et al., 2004). Adipose tissue is a rich source of endocrine factors, collectively termed 'adipokines', which include leptin and adiponectin, and the cytokines, tumour necrosis factor alpha (TNF- $\alpha$ ) and interleukin 6 (IL-6) (Rondinone, 2006). The adipose tissue of obese individuals has also been shown to recruit large numbers of macrophages that secrete a variety of pro-inflammatory signal molecules and cytokines (Fantuzzi, 2005). Obesity is increasingly regarded as low-grade inflammatory condition in which adverse effects are exerted on a variety of target organs including, in all probability, the colon (John et al., 2006).

The plausibility of a role for systemic, low-grade inflammation in the aetiology of sporadic colorectal cancer is supported by the observation that plasma C-reactive protein (CRP) levels have been shown in most studies to be positively correlated with the risk of disease (Otani et al., 2006). One hypothesis is that the pro-inflammatory milieu associated with obesity induces a mucosal signalling cascade involving activation of the transcription factor NF- $\kappa$ B and increased expression of inducible nitric oxide synthase (iNOS) and COX-2 (Gunter et al., 2006). This mechanism is thought to play an important role in the

suppression of apoptosis, which is prerequisite for tumourigenesis (Johnson, 2001).

The abnormally high levels of CRP, TNF- $\alpha$  and IL-6 that occur in the plasma of obese individuals are associated with the development of insulin resistance, a condition characterized by impaired glucose tolerance, elevated plasma levels of insulin and insulin-like growth factor 1 (IGF-1), and with low levels of IGF binding proteins, all of which seem to be associated with heightened vulnerability to colon cancer in human populations (Wei et al., 2005).

Insulin itself causes hyperproliferation of colonic epithelial cells in vivo in the rat and exposure of colorectal cancer cells to insulin, IGF-1 and TNF- $\alpha$  in vitro leads to increased rates of proliferation and impaired apoptosis, both of which favour tumourigenesis (Giovannucci, 2001). Furthermore, recent reports indicate that leptin, which is a secretory product of adipocytes and hence also present at abnormally high levels in the plasma of obese subjects, also promotes mitosis and suppresses apoptosis in colonic epithelial cells (Amemori et al., 2007).

- ***Diet***

A high consumption of fat and red meat is widely regarded as a risk factor for colorectal cancer (Lipkin et al., 1999). Meat consumption is associated with increased intakes of proven rodent carcinogens including heterocyclic amines and polycyclic aromatic hydrocarbons derived from thermal reactions during cooking (Cross and Sinha, 2004). In addition, although the definition of processed meat tends to vary between studies and populations, most

processed meats contain relatively high levels of nitrite and nitrate, which together with protein and haem contribute to the production of mutagenic nitrosamines derived in the faecal stream (Bingham et al., 1996). Red meat is also a rich source of iron, which is itself independently associated with an increased risk of colorectal cancer, and which may act as an intraluminal pro-oxidant (Lund et al., 1999).

The fruit and vegetable intake have anticarcinogenic effects. Apart from studies in which all fruits and vegetables are grouped together, there have been a number of studies in which the roles of specific types of fruit or vegetable have been analysed in depth. The cruciferous vegetables (cabbages, broccoli, sprouts, etc.) have received a great deal of attention, mainly because they contain a particular group of biologically active secondary metabolites, the glucosinolates. There are good evidences for a consistent protective effect of brassica vegetables against colorectal cancer, as well as for cancer of the stomach and lung (van et al., 1999). This issue has become more complex, however, in the light of recent work demonstrating that the protective effects depend upon particular genetic polymorphisms of the glutathione-S-transferase (GST) superfamily. This group of Phase II detoxifying enzymes metabolize environmental chemicals including carcinogens, drugs and phytochemicals. Around half of the individuals in most human populations are homozygous for the null alleles of two genetic subtypes, GSTT1 and GSTM1, and hence cannot express some important components of GST activity (Lin et al., 1998). The GSTM1 null condition might lead to a slower rate of metabolism of glucosinolate breakdown products and enhance the exposure of target tissues to isothiocyanates. Alternatively, GST null phenotypes may be less able to

conjugate the breakdown products in the enterocytes and subsequently export them back into the lumen, thus permitting better absorption of the biologically active constituents (Petri et al., 2003).

Prospective studies showing an adverse effect of meat on the risk of colorectal cancer have also provided evidence for a protective effect of fish (Norat et al., 2005). Although the evidence is still ambiguous, these findings have prompted interest in the possibility that a high dietary intake of n-3 polyunsaturated fatty acids from oily fish may exert anti-carcinogenic effects on the colorectal mucosa, perhaps by reducing the production of pro-inflammatory eicosanoids and inhibiting the expression and activity of COX-2 in a manner analogous to aspirin and other NSAIDs (Hall et al., 2007). Both the epidemiology and the potential mechanisms of action of fatty acids in the aetiology of colorectal cancer have been recently reviewed. The epidemiological evidence is further confused by the observation that the protective effects of fish may be associated with particular genetic polymorphisms affecting the expression of key proteins involved in peroxisome proliferator-activated receptors (PPAR) signalling and COX-2 expression. For example, people carrying a minor polymorphism in PPAR $\alpha$  have an increased risk of adenoma but not tumour, and the protective effects of fish in relation to tumour are more significant in those carrying a particular COX-2 polymorphism. These genes are associated with the control and metabolism of fatty acids and the formation of inflammatory eicosanoids, which supports the idea that it is the particular fatty acid composition of fish that is the key to the protective effects (Siezen et al., 2006).



Folate, which is obtainable only from the diet or from supplements, is essential for normal DNA synthesis and repair because conversion of deoxyuridylate to thymidylate requires the reduction of 5,10-methylenetetrahydrofolate to 5-methyltetrahydrofolate by the enzyme methylenetetrahydrofolate reductase (MTHFR). A deficiency of 5,10-methylenetetrahydrofolate leads to an imbalance in intracellular deoxynucleoside triphosphate pools, misincorporation of uracil into DNA and excess double-strand breaks (Wickramasinghe and Fida, 1994). In principle, low levels of 5-methyltetrahydrofolate can also reduce the availability of the methyl-donor S-adenosyl methionine, causing abnormal methylation of cytosine, both within cytosine-guanine dinucleotide sequences called CpG-islands and associated with the promoter regions of genes, and throughout the non-coding DNA. Abnormal DNA methylation causes aberrant gene expression and genomic instability and this has been proposed as a major cause of genetic damage leading to cancer (Ames, 2001). Folate intake has been shown to be inversely related to the risk of colorectal adenomas in both case-control and prospective studies (Giovannucci et al., 1993).

Vitamin D has also long been thought to be protective against colorectal carcinogenesis, but the situation is complicated because the vitamin D status of an individual depends both upon dietary intake and endogenous synthesis in response to sunlight. The geographical differences in risk of colorectal cancer might be attributable to differences in sunlight exposure. It is difficult to see how this hypothesis can be compatible with the very high incidences of colorectal cancer in Australia and New Zealand, but vitamin D has since been shown to suppress proliferation and promote differentiation in experimental model

systems and a large number of epidemiological studies provide support for the hypothesis that adequate vitamin D status is associated with a moderately reduced risk of colorectal cancer (Garland and Garland, 2006).

### 1.3. Aim

Given the important role of miRNA in gene regulation and in carcinogenesis, we hypothesized that germ-line polymorphisms in the sequence of miRNA or in the 3' UTR regions targeted by miRNAs might alter the strength of miRNA binding, with consequences on regulation of target genes thereby affecting the individual's sporadic colorectal cancer risk. To test this hypothesis, we made a search in dbSNP for the genes coding for all the known human miRNAs (from the microRNA database, miRBase, <http://microrna.sanger.ac.uk>), and we selected 8 polymorphisms of potential functional impact: 7 SNPs identified in 7 pre-miRNA hairpin regions and one in the mature sequence of miR-608.

Moreover, we evaluated 129 genes belonging to different pathways relevant for colorectal carcinogenesis: there are 88 genes involved in inflammatory processes following immune responses, 16 genes involved in the obesity and regulation of insulin levels, and 25 genes that, when mutated, are known to cause genome instability by disrupting the normal DNA repair or the cell cycle checkpoints. We searched for all the polymorphisms residing in the 3' UTRs of the candidate genes, and assessed *in silico* the abilities of each SNP to impact the binding between miRNAs and their target sites.

In order to verify our initial hypothesis we genotyped the 8 SNPs in miRNA sequences and the 15 resulting putatively functional polymorphisms in

697 CRC cases and 624 controls from the Czech Republic, where the incidence of colon cancer ranks the third highest worldwide and is the highest for rectal cancer (Boyle and Langman, 2000; Parkin et al., 2005b).

The SNPs associated with the risk of sporadic colorectal cancer were assayed in a new independent case-control study on three different populations: 377 cases and 326 controls from Barcelona, 224 cases and 449 controls from Florence and Pisa (Italy) and 659 cases and 607 controls from Rhine-Neckar-Odenwald region (Germany), all already collected, promptly available, and already studied for other polymorphisms in previous studies published in international scientific journals.

## **2. MATERIALS AND METHODS**

### **2.1. Selection of genes**

Among the most acknowledged mechanisms which play a role in the etiology of human cancer, we focused on the genes regulating the inflammation, the insulin resistance, and the control of genome stability. Thus, we selected the most studied genes within these categories, according to the information from PubMed ([www.pubmed.com](http://www.pubmed.com)) and CGAP pathways (<http://cgap.nci.nih.gov/Pathways>).

We have selected 129 genes (table 1, in appendix): there are 51 genes involved in inflammatory processes following immune responses, 37 inflammatory genes related to the synthesis of prostaglandins and thromboxanes, 16 genes involved in the regulation of insulin levels and insulin resistance, and 25 genes that, when mutated, are known to cause genome instability by disrupting the normal DNA repair or the cell cycle checkpoints.

For all of them, the 3'UTR regions were selected according to the UCSC genome browser (<http://genome.ucsc.edu>), and defined as transcribed sequences from the stop codon to the end of the last exon of each gene.

### **2.2. Targets prediction**

Putative miRNA binding sites within the 3'UTR of each gene were identified by means of specialized algorithms (e.g. Diana-MicroT, miRBase,

miRanda, PicTar, MicroInspector, and TargetScanS) using the default parameters included in the software, for each of them.

*Diana-MicroT* ([http://www.diana.pcbi.upenn.edu/cgi-bin/micro\\_t.cgi](http://www.diana.pcbi.upenn.edu/cgi-bin/micro_t.cgi)) finds microRNA/target duplexes that are conserved in humans and mice with the minimum free energy (Kiriakidou et al., 2004).

The *miRBase* database (<http://microrna.sanger.ac.uk/targets/v3/>) is divided into three parts: miRBase Registry includes the microRNA gene nomenclature; miRBase Sequence is the primary online repository for miRNA sequence data and annotation; miRBase Targets is a comprehensive new database of predicted miRNA target genes (Griffiths-Jones et al., 2006).

*MiRanda* (<http://www.microrna.org/>) is an algorithm that considers the sequence complementarity between the mature miRNA and the target site, binding energy of the miRNA-target duplex, and the evolutionary conservation of the target position in aligned UTRs of homologous genes (John et al., 2004).

The *MicroInspector* program (<http://mirna.imbb.forth.gr/microinspector/>) generates a list of possible target sites, sorted by free energy values. Adaptation of temperature and free energy setting, followed by a visual inspection of secondary structures allows a detailed analysis. The program uses a “miRNA database” (in multifasta format) based on “the miRNA registry” (<http://www.sanger.ac.uk/Software/rfam/mirna/index.shtml>) (Rusinov et al., 2005).

*TargetScan* (<http://genes.mit.edu/targetscan>) searches the 3'UTRs for segments of perfect Watson-Crick complementarity to bases 2-8 of the miRNA (numbered from the 5' end) and assigns a free energy to miRNA: target site

interaction, given an internal database of miRNA and UTR sequences (Lewis et al., 2003).

*PicTar* (<http://pictar.bio.nyu.edu/>) computes a maximum likelihood score that a given RNA sequence (3' UTR region) is targeted by a fixed set of microRNA. First, PicTar sets the length of putative microRNA binding sites to the length of the corresponding "seed". Second, a short 3'UTR (<300bp) cannot be used to reliably estimate its own background nucleotide frequencies (Krek et al., 2005).

### 2.3. Selection of polymorphisms

The SNPs residing on the miRNA binding sites were found by an extensive search in dbSNP (<http://www.ncbi.nlm.nih.gov/SNP>), using BLAST (<http://www.ncbi.nlm.nih.gov/BLAST/>) and BLAST-SNP algorithms (<http://www.ncbi.nlm.nih.gov/SNP/snpblastByChr.html>).

At present, it is difficult to judge which of the algorithms produces the most reliable and/or sensitive target predictions. The *TargetScan* and *PicTar* algorithms produce similar overall sets of predicted target sites; most of the other algorithms produce results that are significantly different from them. However, in order to obtain comparable measurements of the variation of  $\Delta G$ s between the variant and the common alleles within each polymorphic target site, we used only one algorithm, *miRanda*, which is the most specialized for the calculation of  $\Delta G$ s.

The input list of targets predicted came from all the other algorithms employed, even in the case when one given target was detected by only one

algorithm. The input list of miRNAs in *miRanda* was taken from miRBase, that is the most updated and comprehensive database of miRNAs.

For all the identified SNPs in the miRNA target sites, *miRanda* was run again to assess the Gibbs binding free energy ( $\Delta G$ , expressed in KJ/mol) both for the common and the variant alleles. The difference of the free energies between the two alleles was computed as “variation of  $\Delta G$ ” (i.e.  $\Delta\Delta G$ ). Because in some genes the same sequence is predicted to bind several miRNAs, and thus SNPs in these sequences could impact the binding site of more than one miRNA, we used the sum of all the  $|\Delta\Delta G|_s$  for each SNP ( $|\Delta\Delta G_{tot}|$ ) as parameter for predicting the biological impact of the polymorphism. Although each target can bind only one miRNA at a time in each tissue, this approach is based on the basic idea that the more miRNAs are predicted to bind to a given target, the more likely it is that at least one of them truly binds to the target. Thus, if the algorithms predict that only one miRNA binds to a given target, at least in theory, we should consider this target less important than others predicted to bind many different miRNAs. Parologue miRNAs, when reported to have a different sequence, are considered as different miRNAs. When summing the  $\Delta\Delta G$ s, we did not account for the exactly duplicated miRNA, that are, in this sense, considered as unique.

In order to give a priority list of polymorphisms having real impact on miRNA bindings, we ranked the values of  $|\Delta\Delta G_{tot}|$  and we chose (arbitrarily) the upper tertile of the distribution ( $|\Delta\Delta G_{tot}| \geq 3.10$  KJ/mol) as the significant cut-off. In other words, when the SNP had a  $|\Delta\Delta G_{tot}| \geq 3.10$  KJ/mol was considered biologically relevant, whereas for  $|\Delta\Delta G_{tot}| < 3.10$  KJ/mol the SNPs was considered biologically neutral and not included in further analyses. Moreover,

as second criterion for SNP selection, we excluded the SNPs having the minor allele frequency (MAF) lower than 0.10 (for Caucasians).

Among 129 genes, 51 don't have SNPs in their 3'UTR region, and 29 don't have SNPs in binding sites of miRNA. In the list remain 44 genes and 78 polymorphisms in target sites, but of these only 50 SNPs are validated in the dbSNPs of the site NCBI, and only 32 are a  $MAF \geq 0.10$ . Thus, only polymorphisms falling in the upper tertile of  $|\Delta\Delta G_{tot}|$  with  $MAF \geq 0.10$  were selected for genotyping. Finally, we obtained a list of 15 SNPs (table 2).

Moreover, we made a search in dbSNP for the genes coding for all the known human miRNAs (from the microRNA database, *miRBase*, <http://microrna.sanger.ac.uk>), and we listed 8 polymorphisms of potential functional impact: 7 SNPs identified in pre-miRNA harpin regions and one in the mature sequence of miR-608 (table 3).

## 2.4. Genotyping

### 2.4.1. Study populations

- ***Czech Republic***

A hospital-based case–control study was conducted to assess gene–environment interactions in relation to colorectal cancer risk. Cases were incident patients with a new diagnosis of colorectal cancer attending nine oncological departments (2 in Prague, 1 in Benesov, Brno, Liberec, Ples, Pribram, Usti nad Labem, and Zlin) all over the Czech Republic from September 2004 to February 2006. All cases had histological confirmation of their tumor diagnosis. During the study period, a total of 968 cases were diagnosed with colorectal cancer in these hospitals. This study includes 697 (72%) patients who



could be interviewed and provided biological samples of sufficient quality for genetic analysis. The lost cases were similar to those enrolled with respect to age, sex, tumor location, and extent.

Controls were selected from among patients admitted to the same hospital during the same period and were subjects undergoing colonoscopy for various gastrointestinal complaints. The reasons for undergoing the colonoscopy were: 1) positive Fecal Occult Blood Test (FOBT); 2) haemorrhoids; 3) abdominal pain of unknown origin; 4) macroscopic bleeding. Due to the high incidence of colorectal cancer in the Czech Republic, colonoscopy is largely recommended and practiced. Among 899 selected controls, a total of 624 (70%) were analyzed in this study. Twelve percent could not be interviewed because of refusal or mental or other impairment. Eighteen percent were interviewed but did not provide a blood sample. Controls included people who had benign lower abdominal pains, hemorrhoids with active bleeding, solitary rectal ulcer, lower GI bleeding, diverticular disease, diarrhea, and anemia. People with polyps, adenomas, or other diagnoses related to cancer or to diseases known to predispose to cancer (such as e.g. ulcerative colitis, inflammatory bowel disease, and Crohn's disease) were excluded. Sex and broad age groups were used as stratifying criteria for frequency matching. Both cases and controls had to be in good mental condition, be able to see and hear and follow an interview and (for controls) not have diagnoses clearly related to cancer or chronic inflammatory diseases.

All subjects were informed and gave written consent to participate in the study to allow their biological samples to be genetically analyzed, according to the Helsinki declaration. The design of the study was approved by the Ethical

Committee of the Institute of Experimental Medicine, Prague, Czech Republic (table 4).

- ***Spain***

Cases are patients with a new diagnosis of CRC attending a University Hospital in Barcelona, Spain, between January 1996 and December 1998. All cases had histological confirmation of their tumour diagnosis. During the study period, a total of 523 cases were diagnosed with CRC in the hospital. The study will include those 377 (72%) who could be interviewed and who provided biological samples of sufficient quality for genetic analysis. Refusals were 2%, whereas 14% could not be interviewed because they either had died, had mental or some other impairment, or were released without being approached and could not be traced. Finally, 12% were interviewed but did not provide biological samples. These lost cases were similar to those included with respect to age, sex, tumor location, and extent.

Controls are randomly selected among patients admitted to the same hospital during the same period. To avoid selection bias, the criterion for inclusion of controls was that the reason for the current admittance to the hospital should be a new disease (not previously diagnosed) for that patient. This criterion was used to avoid inclusion of patients with chronic diseases, who might be repeatedly admitted to hospital and modify their habits because of their disease. Sex and broad age groups were used as stratifying criteria for frequency matching.

Both cases and controls had to have good mental condition, and be able to see and hear and follow an interview. From the daily patient admission lists,

candidate controls were approached and, if they met these criteria, they were invited to participate. Among 470 selected controls, a total of 326 (69.4%) were analyzed in this study. Finally, 87 (18.6%) were interviewed but did not provide a blood sample. From a genetic point of view, we consider the hospitalized controls as being representative of the general Spanish population because they came from very different hospital departments and included very different diseases. No restriction criterion was imposed regarding the diagnosis of controls except those previously mentioned. The distribution of controls by diagnostic group was as follows: internal medicine 22%; acute surgery 19%; urology 17%; traumatology 15%; gastroenterology 16%; and circulatory or respiratory 11%. Sixty controls (18%) had a diagnosis of inflammatory conditions that might be related to the studied polymorphisms.

All subjects were informed and gave written consent to participate in the study and to allow their biological samples to be genetically analyzed, according to the Helsinki declaration. The study protocol was cleared by the Ethical Committee of the hospital (table 5).

- ***Italy***

We used 224 cases (patients affected by colorectal cancer) and 226 controls. These samples have been recruited from the hospitals “Careggi” of Florence and “Santa Chiara” of Pisa. Both cases and controls have given the authorization to use their samples of blood for genetic analyses; the approval for the study has been approved from the Ethics Committees.

Cases have been taken consecutively from the units of surgery of the two hospitals. All the subjects who were not admitted to the hospital previously for

chronic conditions were selected. So we selected only cases of sporadic colorectal cancer that they manifested pathology for the first time.

Controls were constituted from donors of blood who voluntarily have joined under consideration. We have excluded all the donors with inferior age to the 39 years, in order to approach the medium age of controls to cases.

- ***Germans***

Colorectal cancer cases and controls were drawn from the German Darmkrebs: Chancen der Verhütung durch Screening (DACHS) study, a large population-based case-control study carried out in the Rhine-Neckar-Odenwald region in the southwest of Germany.

The cases consisted of 1257 unrelated male and female subjects (33-91 years of age; median 68) with incident invasive colorectal cancer diagnosed between January 2003 and March 2005. The median time between diagnosis and ascertainment of the cases was 14 days.

The control group comprised 1307 unrelated male and female individuals (34-94 years of age; median 67). None of the controls had a personal history of colorectal cancer. They were randomly selected from lists of population registries and frequency-matched to cases by 5-year age groups, sex, and county of residence.

Cases and controls were eligible if they were aged 30 years and above, German-speaking, and mentally and physically capable of participating in an in-person interview of about 1 hour. The study was approved by the Ethics Committees of the University of Heidelberg and the State Medical Boards of Baden-Wuerttemberg and Rhineland-Palatinate (Germany) (table 6).

- ***Interviews***

Cases and controls were personally interviewed by trained personnel using a structured questionnaire to determine demographic characteristics and potential risk factors for colorectal cancer. For each subject age and sex were recorded. Study subjects provided information on their lifestyle habits, body mass index (BMI), diabetes, tentative occupational exposure to xenobiotics, and family/personal history of cancer. A detailed dietary history questionnaire focused on average food consumption one year before the diagnosis of disease. Lifelong long-term (at least six consecutive months) drug use was included in the questionnaire. An initial open question was followed by a list of 20 chronic diseases that usually are treated pharmacologically and their treatments were recorded. No drug list was used. For each exposure, the ages at initial use and cessation were recorded and the cumulative duration was computed. Drugs were grouped using the ATC (Anatomical Therapeutic Chemical) classification. Other relevant risk factors explored were smoking, alcohol, physical activity and family history of cancer.

#### **2.4.2. Laboratory Techniques**

- ***TaqMan allelic discrimination***

For 19 polymorphisms, the genotyping was carried out with the 5' nuclease assay (TaqMan; Applied Biosystems, Foster City, CA, USA). Two TaqMan probes were used, one for each allele. Analysis was performed using the ABI PRISM® 7900HT Sequence Detection System and SDS 2.2 software (Applied Biosystems).

The reaction employed the TaqMan Universal PCR Master Mix 2X, TaqMan primers and probes (20X or 40X), water, and 5ng of DNA.

The thermal cycle program used included:

95°C for 10 min (to activate the AmpliTaq Gold enzyme  
Activation)

40 cycles: 95°C for 15 sec (to denature the DNA)  
60°C for 1 min (for the stage of annealing/extension)

The assays ID of probes are:

C\_441625\_10 (for *CD86*, rs17281995);  
C\_7537839\_10 (for *IL12B*, rs1368439);  
C\_1844363\_10 (for *IL16*, rs1131445);  
C\_27465694\_10 (for *NOD2*, rs3135500);  
C\_8356120\_10 (for *INSR*, rs1051690);  
C\_8685434\_10 (for *ALOX15*, rs916055);  
C\_362670\_20 (for *PLA2G2A*, rs11677);  
C\_1311797\_10 (for *Raf1*, rs1051208);  
C\_2184704\_10 (for *CD4*, rs3213428);  
C\_15831555\_10 (for *INSR*, rs2860171);  
C\_9924799\_10 (for *miR-605*, rs2043556)  
C\_15880380\_10 (for *miR-492*, rs2289030);  
C\_2826025\_10 (for *miR-608*, rs4919510)  
C\_15946974\_10 (for *miR-146a*, rs2910164);  
C\_31185852\_10 (for *miR-196a-2*, rs11614913);

The SNPs rs16870224 (*PTGER4*), rs11515 (*CDKN2A*), rs2973016 (*GHR*), rs17759989 (for *miR-633*) was genotyped according to an ABI assay-by-design.

The sequences of primers and probes for these polymorphisms are:

*PTGER*: 5'-AGTGTCTCACTAAAGCATGAAATGTGAA-3' (forward primer),  
5'-CCTTGTATTAACAATAAACCTTCTCACAGAGA-3' (reverse primer),  
VIC-5'-CATACGATTTAAGGTATTTAA-3' (reporter1)  
FAM-5'-CATACGATTTAAGTATTTAA-3' (reporter2)

*CDKN2A*: 5'-CCCCGATTGAAAGAACCAGAGA-3' (forward primer)  
5'-AGGACCTTCGGTGACTGATGAT-3' (reverse primer)  
VIC-5'-CTAAGTTTCCCGAGGTTT-3' (reporter1)  
FAM-5'-CTAAGTTTCCCGAGGTTT-3' (reporter2)

*GHR*: 5'-ACTCGGCCTCCCAAAGTG-3' (forward primer)  
5'-TCCCAATGCTCTCAAGAATGATGTT-3' (reverse primer)  
VIC-5'-TAAGCCACCGTGCCCA-3' (reporter1)  
FAM-5'-AGCCACCATGCCCA-3' (reporter2)

*miR-633*: 5'-CTCAACCTCTCTTAGCCTCTGTTTC-3' (forward primer)  
5'-GTGGTAGATACTATTAGCCTTCTCATTTTAGG-3' (reverse primer)  
VIC-5'-TTTATTGCGGTAGATACTAT-3' (reporter1)  
FAM-5'-TTATTGCGGTAGGTACTAT-3' (reporter2)

- ***Oligonucleotide allele specific PCR (ASO-PCR)***

For 2 polymorphisms (rs360727, *IL 18* and rs3746444, *miR-499*) genotyping was carried out with the oligonucleotide allelic specific PCR.

In this technique three oligonucleotide primers are used: a primer forward (or reverse, F1 or R1) with a sequence complementary to the interest DNA fragment that finishes at the level of the allele wild-type of the polymorphism, an other primer forward (or reverse, F2 or R2) that, instead, it recognizes the variant base present on the polymorphic site, and common primer reverse (or forward, R or F).

The sequences of primers for these polymorphisms are:

*IL 18* (rs360727, G/A):      R1: 5'-CGGCTCACCACAACCTCC-3'  
    R2: 5'-CGGCTCACCACAACCTCT-3'  
    F: 5'-TGTTCACTGTTCAAAACGAAGAC-3'

SIZE PCR: 273bp

*miR-499* (rs3746444, T/C):      R1: 5'-GTTTAACTCCTCTCCACGTGAACA-3'  
    R2: 5'-GTTTAACTCCTCTCCACGTGAACG-3'  
    F: 5'-CCAGAATTGGATGCCGCAGT-3'

SIZE PCR: 267 bp

We prepared two different mixes, since in one the primers are added the primers F and R1 in order to amplify the region with the allele wild-type, in the other are added the primers F and R2 in order to amplify the region with the variant allele.



For the SNP within the gene *IL18* the protocol is:

**MIX1** (for the wild-type allele)

Primers (10 µM): 0.5 µl F + 0.5 µl R1

Buffer BD (10X): 2 µl

MgCl<sub>2</sub> (25mM): 1.2 µl

dNTPs (2mM): 2 µl

H<sub>2</sub>O: 10,6 µl

Hotfire® Taq polimerasi: 0,2 µl

DNA: 3 µl (30 ng/ µl)

**MIX2** (for the variant allele)

Primers (10 µM): 0.5 µl F + 0.5 µl R2

Buffer BD (10X): 2 µl

MgCl<sub>2</sub> (25mM): 1.2 µl

dNTPs (2mM): 2 µl

H<sub>2</sub>O: 10,6 µl

Hotfire® Taq polimerasi: 0,2 µl

DNA: 3 µl (30 ng/ µl)

For the SNP within *miR-499* the protocol is:

**MIX1** (for the wild-type allele)

Primers (10 µM): 1 µl F + 1 µl R1

Buffer BD (10X): 2 µl

MgCl<sub>2</sub> (25mM): 0,6 µl

dNTPs (2mM): 2 µl

H<sub>2</sub>O: 10,2 µl

Hotfire® Taq polimerasi: 0,2 µl

DNA: 3 µl (30 ng/ µl)

**MIX2** (for the variant allele)

Primers (10 µM): 1 µl F + 1 µl R2

Buffer BD (10X): 2 µl

MgCl<sub>2</sub> (25mM): 0,6 µl

dNTPs (2mM): 2 µl

H<sub>2</sub>O: 10,2 µl

Hotfire® Taq polimerasi: 0,2 µl

DNA: 3 µl (30 ng/ µl)

The thermal cycle program used for both SNPs included:

	95 °C for 8 min (to active the Hotfire® Taq polim erasi)
20 cycles	95°C for 30 sec (to denature the DNA)
	68°C for 30 sec, -1°C/cycles (hybridization of pr imers)
	72°C for 1 min and 30 sec (the stage of extension )
20 cycles	95°C for 30 sec (to denature the DNA)
	51°C for 30 sec (hybridization of primers)
	72°C for 1 min and 30 sec (the stage of extension )
	72°C for 10 min (to complete the synthesis)
	4°C forever

The visualization of the PCR products was carried out through electrophoresis in 2% agarose gel, obtained through the polymerization of 5 g of agar in 250 ml TBE 0,5X (TRIS borat-EDTA buffer) , with the addition of 8 µl of Ethidium Bromide (10mg/ml), at 180V for 20 minutes.

- ***Fragment Analysis***

For 2 polymorphisms (rs3783553, *IL1A* and rs5745925, *miR-631*) genotyping was carried out with the DNA Fragment Analysis. In this methodical the regions containing the SNPs are amplified by duplex PCR using fluorescently labeled reverse and unlabeled forward primers. The PCR amplicons are separated by size using capillar electrophoresis; then the dye labeled products are identified by fluorescence detection. Analysis was performed using the ABI PRISM® 3100 Genetic Analyzer and the

GeneMapper® Version 3.7 Software. Before the use the DNA samples were put in plates from 96 wells.

The sequences of primers for the polymorphisms are:

*IL1A*

(rs3783553, TTCA/-): F: 5'-CTCTGAGTGTGACCAGGCATC-3'  
R: Fam- 5'-GAACCTAGAAGAAACATTGATTACA- 3'

For the wild type allele the size PCR is: 363 bp

For the wild type variant allele the size PCR is: 359 bp

*MiR-631*

(rs5745925, CT/-): F: 5'-GCTGAGGTCACATCACCAGA -3'  
R: Fam-5'-TCTCTGACCCGTAGCCTTTG -3'

For the wild type allele the size PCR is: 176 bp

For the wild type variant allele the size PCR is: 174 bp

The protocol of duplex PCR for plate is:

Primers(10 µM): 100µl F<sub>(IL1A)</sub>+ 100µl R<sub>(IL1A)</sub>+100µl F<sub>(miR-631)</sub>+100µl R<sub>(miR-631)</sub>

Buffer BD (10X): 200 µl

MgCl<sub>2</sub> (25mM): 240 µl

dNTPs (2mM): 200 µl

H<sub>2</sub>O: 630 µl

Hotfire® Taq polimerasi: 30 µl

DNA: 3 µl (30 ng/ µl)

The thermal cycle program used included:

35 cycles     95°C for 1 min (to denature the DNA)  
                  58°C for 30 sec (hybridization of primers)  
                  72°C for 10 min (the stage of extension)  
  
                  72°C for 10 min (to complete the synthesis)  
                  4°C forever

After the amplification, to 1 µl of PCR product for everyone samples were added 10 µl of Rox500 and HiDi Formamide (dilution 1:40).

### **2. 4. 3. Statistical Analysis**

Each polymorphism was tested to ensure the fitting with Hardy–Weinberg equilibrium with alpha-threshold of 0.05. We used a multivariate logistic regression analysis to examine associations between these polymorphisms and CRC risk by estimating the odd ratios (ORs) and 95% confidence intervals (CIs), adjusted for sex and age (as linear variable) as covariates. Genotypes have been divided into three groups: major allele homozygous (also defined as +/+), heterozygous (+/-), and homozygous variant (-/-). This analysis had been done using a co-dominant model: ORs and 95% CI were calculated for each genotype compared to the group +/+. Test for linear trend of ORs were calculated using the categorized variable as quantitative, after assigning a linear score to each ordered category. When associations were detected, also the recessive (homozygotes variant vs heterozygotes+homozygotes common allele) and dominant (homozygotes variant+heterozygotes vs homozygotes common allele) models were evaluated. All statistical analyses were conducted using STATA software version 8.0 (Stata Corp. LP, College Station, TX).

### 3. RESULTS

Among the 129 starting candidate genes, 56 did not present any SNP in their 3'UTR regions and 29 genes did not show SNPs in miRNA target sequences. The remaining 44 genes showed 78 SNPs in the predicted miRNA binding sites. However, among them, only 15 SNPs were positive for both the selection criteria ( $|\Delta\Delta G_{\text{tot}}| \geq 3.10$  KJ/mol and  $\text{MAF} > 0.10$ ) and were further investigated in the case-control study.

Table 2 shows the nucleotide variation, the predicted binding miRNAs, variation of  $\Delta G$  ( $|\Delta\Delta G|$ ) for each miRNA, and the overall  $|\Delta\Delta G_{\text{tot}}|$  for the 15 selected SNPs. These SNPs were in the 3' UTR of *CD4*, *CD86* and *NOD2* (receptors involved in the immune response), *IL-1A*, *IL-18*, *IL16* and *IL12B* (cytokines), *ALOX15*, *PLA2G2A*, and *PTGER4* (bio-synthesis of prostaglandins and thromboxanes), *INSR* and *GHR* (obesity and insulin resistance), and *RAF-1* and *CDKN2A* (early and late stages of the tumour). To this list we added 7 SNPs in the sequence of pre-miRNA (miR-146a, miR-499, miR-605, miR-492, miR-631, miR-633, miR-196) and one in the mature sequence of miR-608. The variations and the frequencies of these polymorphisms are in appendix (table 3).

We tested the genotypes for Hardy-Weinberg equilibrium, using  $p=0.05$  as significance threshold and all 23 SNPs were in equilibrium in controls. Table 7 shows the statistical analyses, for the co-dominant model only, for the Czech Republic population. Both the heterozygotes and homozygotes for the variant allele rs17281995 within *CD86* were associated with an increased risk of CRC

(OR=1.38; 95%CI=1.05-1.83 and OR= 2.74; 95%CI= 1.24-6.04, respectively). Moreover, the homozygotes for the variant allele rs1051690 within *INSR* showed an OR of 1.94 with a 95% Confidence Interval of 1.03-3.66. It is important to note that the association found for *CD86* in the dominant model, with an OR of 1.48 (95%CI= 1.13-1.94; P= 0.015) was also very close to this tight limit of significance. Stratifying for the site of cancer, the increased risk for *CD86* was observed both for the colon and rectum; however, it was more evident in the rectum (OR=1.5; 95%CI 1.05-2.15; p=0.028, dominant model).

For these two polymorphisms we used a multivariate logistic regression analysis with the other populations: in table 8 there are the results for the co-dominant model for the Spain population. For the SNP in *CD86* the association for homozygotes for the variant allele was confirmed, but the 95% Confidence Interval was too wide (OR=1.34; 95%CI 0.51-3.48). The heterozygotes was no more associated with an increased risk of CRC (OR=0.97; 95%CI 0.66 1.42). Nevertheless the homozygotes for the variant allele in *INSR* showed an OR of 3.20 with a 95% Confidence Interval of 1.05-9.78, in agreement with the previous results on the Czech Republic population. With the recessive model the association was confirmed for both SNPs (OR=1.35; 95%CI=0.55-3.49 and OR= 3.20; 95%CI= 1.05-9.75, respectively). Also in this case the 95% Confidence Interval was too wide for *CD86*.

Table 9 show the co-dominant model for *CD86* and *INSR* in the Italian population. The odd ration for the homozygotes for the variant allele rs17281995 within *CD86* was 0.30 and 95%CI=0.4-2.32, whereas for the heterozygotes the OR was 1.18 with a 95%CI= 0.68-2.02. The homozygotes within *INSR* show an OR of 1.74 (95%CI= 0.29-10.48).

These discordant results with the first population could be due to a reduced number of samples used for the study.

In table 10 there are the results for the German population. For the homozygotes for the variant allele of *CD86* the OR was 1.01 (95%CI= 0.60-1.69), while for the heterozygotes the OR was 1.15 (95%CI= 0.95-1.37). The heterozygotes for *INSR* show a OR of 1.1 (95%CI= 0.90-1.30).

Finally, we made a general multivariate logistic regression analysis using all the samples (Czech, Spanish, Italian, and German) (table 11). We confirmed the association between risk of CRC and the polymorphisms in *CD86* (rs17281995) (OR=1.40, 95%CI= 1.02-2.01 and OR= 1.15, 95%CI=1.01-1.31, respectively for the homozygotes and the heterozygotes, co-dominant model) and *INSR* (rs1051690) (OR=1.41, 95%CI= 1.01-1.94, for the heterozygotes, co-dominant model).

## 4. DISCUSSION

In the present study two polymorphisms (rs17281995 in *CD86* and rs1051690 in *INSR*) were found associated with an increased risk of CRC. We were aware that there could be a potential bias in this study because controls were selected among people undergoing colonoscopy. However, in terms of diet, socio-economic status, ethnic and familial origins, and all the variables collected with in questionnaires, there is no evidence that the controls are not representative of the general population. Five different miRNAs bind to a target site that contains the same polymorphism within *CD86*: miR-337, miR-582, miR-200a\*, miR-184, and miR-212 (table 2). When C substitutes G, miR-337, miR-582, and miR-200a\* are predicted to bind less tightly to the *CD86* 3'UTR. In contrast, miR-184 and miR-212 increase their binding affinity. Among all the SNPs analyzed, this is the only one that impacts the binding site of 5 different miRNAs at the same time. In other words, the SNP that produces the largest predicted effect it is also the most strongly associated with risk in the present case-control association study. Furthermore, it should be considered that part of the selected miRNA target sites could have been the result of erroneous prediction by the algorithms (limited experimental information is available for most of the selected miRNAs, according to the latest update of Tarbase 4.0; <http://www.diana.pcbi.upenn.edu/tarbase.html>).

It should be stressed that miR-582 was shown to be present in normal colon tissues, and its expression appeared to be reduced in CRC following measurements done with quantitative real-time PCR (Cummins et al., 2006).



We could speculate that this finding seems in agreement with the fact that the SNP predicted to cause a reduced binding of miR-582 to its target is associated with increased risks of CRC. Taken together, all these facts encourage further investigation, by replicating the results in other populations and by making in vitro experiments in the biological activity of these polymorphisms.

CD86, with CD80, is a co-stimulatory ligand expressed on the surface of the antigen presenting cells (dendritic cells, macrophages, and B cells), in the immune system (Orabona et al., 2004). Binding of these molecules to the T cell co-stimulatory receptors, CD28 and CTLA-4, is essential for the activation and regulation of T-cell immunity (Bhatia et al., 2006). Functionally, the binding of CD28 with CD80 and CD86 ligands delivers a positive signal to T-cells that culminates in T-cell proliferation, cytokine production, and prevents induction of T-cell tolerance (Linsley et al., 1991). In contrast, the interaction of CTLA-4 with CD80/CD86 results in negative signaling leading to attenuation of T-cell activation. Such processes are essential for the regulation of a T-cell response and maintaining T-cell homeostasis (Tivol et al., 1995). CD80 and CD86 have been shown to modulate also the Th1/Th2 T-cell differentiation. While CD80 preferentially favours Th1 type T-cell differentiation, CD86 augments IL-4 production and overall Th2 type T-cell responses (mostly producing anti-inflammatory cytokines) (Suvas et al., 2002). Normal colonic tissue expresses CD80 and CD86 only sporadically, whereas in the pathogenesis of Crohn's and inflammatory bowel diseases, a sharp increase is observed paralleling the impaired and/or excessive activation of mucosal macrophages and T-lymphocytes (Vuckovic et al., 2001). It should be considered that even a normal colonic mucosa undergoes a continuous state of sub-chronic inflammation due

to the normal bacterial activity present within its lumen (Kelly and Conway, 2005). Thus, our results support the hypothesis that the regulation of the immune response and its consequent level of local inflammatory micro-environment could play an important role for the modulation of the individual risk. CD86 is likely one of the key molecules involved in these processes. If we assume the role of miR-582 for CD86, one could speculate that the rare variant of rs17281995 is linked to increased levels of CD86, thereby favoring an inflammatory micro-environment predisposing to CRC.

Although less strongly associated with the risk of CRC, it is worth to note that also the variation G/A within the 3' UTR of *INSR* has a predicted strong impact in the biology of the miRNA target, as it can be observed by the fifth strongest  $|\Delta\Delta G_{tot}|$  (11.25 KJ/mol, see table 2). It is important to stress that this value is due largely by the predicted alteration with miR-612 (7.41 KJ/mol), the second strongest individual  $|\Delta\Delta G|$  observed in this study.

The levels of glucose and insulin increase after each meal. Following binding of insulin to the extra-cellular portion of the insulin receptors (IRs), a second messenger system diverges into two separate pathways that regulate distinct biological effects: the PI3K pathway or the MAP kinase pathway (Taniguchi et al., 2006). In insulin resistance, the ability of insulin to initiate these phosphorylation cascades is diminished (Pirola et al., 2004). Epidemiologic studies indicates that the insulin resistance, elevated fasting plasma insulin, glucose and free fatty acids, glucose intolerance, increased body mass index, and visceral adiposity, are all associated with adenomas, a precursor lesion for CRC (Keku et al., 2005). Several mechanisms exist whereby insulin signaling can be modulated at the level of the *INSR*. Generally,

these processes involve either the modification of *INSR* contained in the cells at the level of transcription or protein degradation, or by modification of the enzymatic activity of individual *INSR* (Youngren, 2007). Puig and Tjian (Puig and Tjian, 2005) have demonstrated that the FOXO1 transcription factor mediates a down-regulation of *INSR* transcription in response to insulin stimulation of IR signaling (Okabayashi et al., 1989b). Insulin also reduces the effective concentration of cellular IR by increasing internalization and degradation of the protein as well (Okabayashi et al., 1989a). Activation of this system produces a down-regulation of multiple components of the insulin signaling system (Rome et al., 2004). The SNP rs1051690 3'UTR could affect *INSR* regulation by a miRNA mediated mechanism. To confirm that the regulation of transcription or translation of *INSR* is an important mechanism implicated in colorectal risk, there is a study of Pechlivanis and your group. They found a significant association between a *INSR* (A-603G, rs1864010) genotypes and risk of CRC (global  $\chi^2=6.15$ ,  $P=0.046$ ) (Pechlivanis et al., 2007). In the *INSR* gene, the region 1800 bp upstream of the translation start site contains many regulatory elements which have been reported to be required for proper expression of *INSR* and appropriate insulin signaling and action (Foti et al., 2003). The removal or mutational inactivation of the four Sp1 elements present between -618 and -593 bp has been shown to reduce promoter activity of the gene by about 90% (Yoshizato et al., 2001). The A-603G polymorphism lies within the consensus sequence for Sp1 binding and has been shown to affect Sp1 binding. However, a later study has not observed any effect of this SNP on the promoter activity (Haruta et al., 1995). Interestingly, in the G allele carrier status was associated with a significantly decreased risk for CRC. In

consistence with these results, the A allele has been suggested to be a risk factor for insulin resistance (Kadowaki et al., 1990).

## 5. CONCLUSIONS

In conclusion, we have analyzed 129 CRC-related genes and selected with *in silico* methods 15 SNPs predicted to have an impact on miRNA binding. To this list we added 7 SNPs in pre-miRNA sequences and 1 in a mature sequence of miR-608. Two of them were associated with CRC risk. This study provides evidence that these SNPs in microRNA binding sites may be important for cancer risk and supports future work to validate the results in other well characterized populations as well as to explore the biological significance of these particular SNPs. Our work is the first showing an association between cancer risk and polymorphic miRNA binding sites. These results show also that a careful *a priori* selection of candidate SNPs could increase the signal-to-noise ratio in case-control association studies. The proposed approach could ease the identification of functionally relevant polymorphisms involved in carcinogenesis to be further analysed by appropriate (but time consuming) experiments of molecular biology.

## 6. FUTURE PROSPECTIVES

For the future we will follow two lines of research. First, we will continue the association study on a new set of genes. In particular, we will apply the approach to 144 candidate cancer genes selected by Vogelstein. In this study, 20857 transcripts were analysed from 18191 human genes, including the great majority of those that encode proteins. The genes were sequenced in 11 colorectal cancers. Any gene that was mutated in the tumour but not in normal tissue from the same patient was analysed in 24 additional tumors. Selected genes were further analyzed in another 96 colorectal cancer to better define their mutation frequency (Wood et al., 2007).

Second, we will study, with the use of gene reporters, the SNPs in *CD86* and *INSR*. We will clone the 3'-UTR of the target gene downstream of the luciferase (with the wild type allele and with the variant allele in the polymorphic site) and will transfect these recombinant plasmids and a miRNA of interest into a host cell. Thus, Lucifer's activity will be measured.

## APPENDIX

Table 1. Starting list of candidate genes evaluated for the presence of polymorphic miRNA target sites.

<i>Gene symbol</i>	<i>Gene name</i>	<i>Gene symbol</i>	<i>Gene name</i>
<b>ACDC</b>	adiponectin and collagen domain containing	<b>INS</b>	insulin
<b>ALOX12</b>	arachidonate 12-lipoxygenase	<b>INSR</b>	insulin receptor
<b>ALOX12B</b>	arachidonate 12-lipoxygenase, 12R type	<b>IRS1</b>	insulin receptor substrate 1
<b>ALOX15</b>	arachidonate 15-lipoxygenase	<b>IRS2</b>	insulin receptor substrate 2
<b>ALOX15B</b>	arachidonate 15-lipoxygenase, second type	<b>JAK2</b>	Janus kinase 2
<b>ALOX5</b>	arachidonate 5-lipoxygenase	<b>KRAS</b>	v-Ki-ras2 Kirsten rat sarc. viral onc. homolog
<b>APC</b>	Adenomatosis polyposis coli	<b>LEP</b>	leptin
<b>AXIN1</b>	Axin	<b>LEPR</b>	leptin receptor
<b>BTRC</b>	beta-transducin repeat-containing protein	<b>MCC</b>	mutated in colorectal cancers
<b>CD28</b>	CD28 antigen	<b>MDM2</b>	Mdm2, transf. 3T3 cell double minute 2
<b>CD3D</b>	CD3D antigen, delta polypeptide	<b>MLH1</b>	mutL hom. 1, colon canc, nonpolyp. type 2
<b>CD3E</b>	CD3E antigen, epsilon polypeptide	<b>MSH2</b>	mutS hom. 2, colon canc, nonpolyp. type 1
<b>CD3G</b>	CD3G antigen, gamma polypeptide	<b>MSH6</b>	mutS homolog 6 (E. coli)
<b>CD3Z</b>	CD3Z antigen, zeta polypeptide	<b>NFkB1</b>	Nucl. fact. K-light polyp. enhancer B-cells
<b>CD4</b>	CD4 antigen (p55)	<b>NOD1</b>	Nucleotide-binding oligomer. domain1
<b>CD80</b>	CD80 antigen	<b>NOD2</b>	Nucleotide-binding oligomer. domain2
<b>CD86</b>	CD86 antigen	<b>NR1H2</b>	Nucl. recept. subfam. 1, group H, member 2
<b>CDKN2A</b>	p16(INK4); p14(ARF)	<b>NR1H3</b>	Nucl. recept. subfam. 1, group H, member 3
<b>CTBP1</b>	c-terminal-binding protein 1	<b>NRAS</b>	neuroblast. RAS viral (v-ras) onc. homolog
<b>CTLA4</b>	Cytotoxic T-lymph.-associated protein 4	<b>PGDS</b>	prostaglandin D2 synthase
<b>CTNNB1</b>	Catenin (cadherin-associated protein), beta	<b>PI3K</b>	phosphoinositide-3-kinase, gamma pol.
<b>DCC</b>	deleted in colorectal carcinoma	<b>PLA2G10</b>	phospholipase A2, group X
<b>DVL2</b>	Dishevelled, dsh homolog 2 (Drosophila)	<b>PLA2G2A</b>	phospholipase A2, group IIA
<b>ELOVL6</b>	long chain fatty acyl elongase	<b>PLA2G4A</b>	phospholipase A2, group IVA
<b>FABP5</b>	fatty acid binding protein 5	<b>PLA2G6</b>	phospholipase A2, group VI
<b>FADS1</b>	fatty acid desaturase 1	<b>PMS1</b>	PMS1 postmeiotic segregation increased 1
<b>FADS2</b>	fatty acid desaturase 2	<b>PMS2</b>	PMS2 postmeiotic segregation increased 2
<b>FADS3</b>	fatty acid desaturase 3	<b>PPAR</b>	peroxisome prolifer. activated receptor, delta
<b>FZD1</b>	Frizzled homolog 1 (Drosophila)	<b>PPARG</b>	peroxisome prolifer. activated receptor, gamma
<b>GH</b>	growth hormone	<b>PPARGC1A</b>	PPARG co-activator 1, alpha
<b>GHR</b>	growth hormone receptor	<b>PRKCB1</b>	Protein kinase C, beta 1
<b>GMCSF</b>	Colony stimulating factor 2	<b>PTGDS</b>	prostaglandin D2 synthase, brain
<b>AES</b>	GROUCHO/amino-terminal enhancer of split	<b>PTGER1</b>	prostaglandin E receptor 1, EP1
<b>GSK3B</b>	Glycogen synthase kinase 3 beta	<b>PTGER2</b>	prostaglandin E receptor 2, EP2
<b>HNF4A</b>	hepatocyte nuclear factor 4, alpha	<b>PTGER3</b>	prostaglandin E receptor 3, EP3
<b>HRAS</b>	v-Ha-ras Harvey rat sarc. viral onc. homolog	<b>PTGER4</b>	prostaglandin E receptor 4, EP4
<b>IFNG</b>	Interferon gamma	<b>PTGES</b>	prostaglandin E synthase
<b>IGF1</b>	Somatomedin C	<b>PTGES2</b>	prostaglandin E synthase 2
<b>IGF1R</b>	insulin-like growth factor 1 receptor	<b>PTGIS</b>	prostaglandin I2 synthase
<b>IGF2</b>	Somatomedin A	<b>PTGS1</b>	Cyclooxygenase-1
<b>IGFALS</b>	IGF, acid labile subunit	<b>PTGS2</b>	Cyclooxygenase-2
<b>IGFBP1</b>	insulin-like growth factor binding prot. 1	<b>RAF1</b>	v-raf-1 murine leuk viral onc. homolog 1
<b>IGFBP3</b>	insulin-like growth factor binding prot. 3	<b>RARA</b>	retinoic acid receptor, alpha
<b>IL10</b>	Interleukin 10	<b>RXRB</b>	retinoic X receptor, beta
<b>IL12A</b>	Interleukin 12	<b>SCD</b>	stearyl-CoA desaturase
<b>IL12B</b>	Interleukin 12B	<b>SMAD4</b>	mothers against decapentaplegic homolog 4
<b>IL12RB2</b>	Interleukin 12 receptor, beta 2	<b>SREBF1</b>	sterol reg. element binding transc. fact. 1
<b>IL15</b>	Interleukin 15	<b>STAT4</b>	Signal transducer/activator of transcr. 4
<b>IL16</b>	Interleukin 16	<b>TBXAS1</b>	thromboxane A synthase 1
<b>IL18</b>	Interleukin 18	<b>TCRA</b>	T cell receptor alpha locus
<b>IL1A</b>	Interleukin 1, alpha	<b>TCRB</b>	T cell receptor beta locus
<b>IL1B</b>	Interleukin 1, beta	<b>TGFB1</b>	transforming growth factor beta
<b>IL1R1</b>	interleukin 1 receptor, type I	<b>TLR1</b>	<i>Toll-like receptor-1</i>
<b>IL1R2</b>	interleukin 1 receptor, type II	<b>TLR2</b>	<i>Toll-like receptor-2</i>
<b>IL2</b>	Interleukin 2	<b>TLR3</b>	<i>Toll-like receptor-3</i>
<b>IL23</b>	Interleukin 23	<b>TLR4</b>	<i>Toll-like receptor-4</i>
<b>IL2RA</b>	Interleukin 2 receptor, alpha	<b>TLR5</b>	<i>Toll-like receptor-5</i>
<b>IL2RB</b>	Interleukin 2 receptor, beta	<b>TLR6</b>	<i>Toll-like receptor-6</i>
<b>IL2RG</b>	Interleukin 2 receptor, gamma	<b>TLR7</b>	<i>Toll-like receptor-7</i>
<b>IL3</b>	Interleukin 3	<b>TLR8</b>	<i>Toll-like receptor-8</i>
<b>IL4</b>	Interleukin 4	<b>TLR9</b>	<i>Toll-like receptor-9</i>
<b>IL5</b>	Interleukin 5	<b>TNFA</b>	Tumor necrosis factor alpha
<b>IL6</b>	Interleukin 6	<b>TP53</b>	tumor protein p53 (Li-Fraumeni syndrome)
<b>IL7</b>	Interleukin 7	<b>WNT1</b>	Wingless-type MMTV integration site fam. 1
<b>IL8</b>	Interleukin 8		

*Table2.* Candidate SNPs in miRNA target sequences, ranked by  $\Delta\Delta G$ , expected to have a high impact on the  $\Delta G$  ( $|\Delta\Delta G_{TOT}| > 3.10$  KJ/mol) and with a frequency higher than 0.10 in Caucasians.

<i>Gene</i>	<i>dbSNP ID</i>	<i>Variation</i>	<i>Frequency</i>	<i>miRNA</i>	<i>/<math>\Delta\Delta G</math>/</i> KJ/mol	<i>/<math>\Delta\Delta G_{TOT}</math>/</i> KJ/mol
<i>ALOX15</i>	rs916055	T/C	0,27	miR-588	4,29	7,50
				miR-183	3,21	
<i>CD4</i>	rs3213428	A/G	0,21	miR-518	3,10	3,10
<i>CD86</i>	rs17281995	G/C	0,11	miR-337	5,76	20,81
				miR-184	4,51	
				miR-200a*	4,32	
				miR-212	2,20	
				miR-582	4,02	
<i>CDKN2A</i>	rs11515	C/G	0,18	miR-601	5,27	5,27
<i>GHR</i>	rs2973016	G/A	0,11	miR-608	2,85	8,52
				miR-196b	5,67	
<i>IL12B</i>	rs1368439	A/C	0,32	miR-513	6,50	15,10
				miR-210	4,20	
				miR-27a	2,20	
				miR-27b	2,20	
<i>IL16</i>	rs1131445	T/C	0,31	miR-135a	3,93	9,53
				miR-135b	3,93	
				miR-143	0,66	
				miR-18	0,51	
				miR-18a	0,50	
<i>IL18</i>	rs360727	G/A	0,17	miR-197	6,99	11,39
				miR-361	4,40	
<i>IL1A</i>	rs3783553	TTCA/-	0,41	miR-180	19,91	58,35
				miR-422b	0,52	
				miR-206	13,85	
				miR-608	24,07	
<i>INSR</i>	rs1051690	G/A	0,15	miR-612	7,41	11,25
				miR-618	3,84	
<i>INSR</i>	rs2860171	C/T	0,16	miR-608	7,12	15,04
				miR-526a	7,92	
<i>NOD2</i>	rs3135500	G/A	0,33	miR-158	6,87	14,69
				miR-215	6,87	
				miR-98	0,64	
				miR-573	0,31	
<i>PLA2G2A</i>	rs11677	C/T	0,11	miR-187	7,90	14,09
				miR-638	2,81	
				miR-154	1,81	
				miR-453	1,37	
				miR-296	0,20	
<i>PTGER4</i>	rs16870224	G/A	0,10	miR-9	4,40	10,30
				miR-30a-3p	3,10	
<i>RAF1</i>	rs1051208	C/T	0,22	miR-213	3,98	5,28



Table3. Candidate SNPs in sequences encoding miRNAs

<b>Gene</b>	<b>dbSNP</b>	<b>Variation</b>	<b>Frequency</b>
<i>miR-608</i>	rs4919510	C/G	0.167
<i>miR-146a</i>	rs2910164	G/C	0.239
<i>miR-499</i>	rs3746444	A/G	0.174
<i>miR-605</i>	rs2043556	U/C	0.139
<i>miR-492</i>	rs2289030	C/G	0.082
<i>miR-631</i>	rs5745925	CT/-	0.065
<i>miR-633</i>	rs17759989	A/G	0.042
<i>miR-196a-2</i>	rs11614913	C/T	0.440

Table4. Characteristics of cases and control subjects: Czech Republic population.

		<b>Cases (n=697)</b>	<b>Controls (n=624)</b>
<b>Males/Females (%)</b>		57.2/48.2	53.6/46.4
<b>Age</b>	<b>Average Lower Quartile Upper quartile</b>	61.1 yrs 55 yrs 69 yrs	56.0 yrs 47 yrs 66 yrs
<b>Rectal cancer (%) Colon cancer (%)</b>		38.4 61.7	
<b>Smoking</b>	<b>Non smokers (%) Ex-smokers (%) Smokers (%)</b>	53.6 32.3 14.1	52.8 26.4 20.9
<b>Positive family history of colorectal cancer</b>	<b>no (%) yes (%)</b>	82.0 18.0	84.3 15.7
<b>Percentage of strictly vegetarian (%)</b>		10.9	7.8
<b>Alcohol consumption</b>	<b>no (%) yes (%)</b>	47.3 52.7	40.2 59.8
<b>Mean gr of alcohol/day</b>		23.1	22.9
<b>Living Place</b>	<b>City (%) Suburbs (%) Country (%)</b>	57.2 15.5 27.3	56.2 21.2 22.7
<b>Education</b>	<b>Basic (%) High School (%) University (%)</b>	33.5 51.6 14.9	25.4 54.8 19.8
<b>Mean BMI (Kg/m<sup>2</sup>)±SD</b>		26.7±4.3	26.7±4.5
<b>Distribution of the populations according to BMI:</b>	<b>&lt;18.5 Kg/m<sup>2</sup> (%) 18.5-24.9 Kg/m<sup>2</sup> (%) 25-29.9 Kg/m<sup>2</sup> (%) 30-34.9 Kg/m<sup>2</sup> (%) ≥35 Kg/m<sup>2</sup> (%)</b>	1.56 37.0 42.5 15.1 3.80	0.73 37.9 41.5 15.3 4.61

Table5. Characteristics of cases and control subjects: Spain population.

		<b>Cases</b> (n=337)	<b>Controls</b> (n=326)
<b>Males/Females (%)</b>		59,7/40,3	53.1/46.9
<b>Age</b>	<b>&lt;58</b>	21,7	26,4
	<b>58-67</b>	24,7	26,7
	<b>68-75</b>	29,2	20,2
	<b>&gt;75</b>	24,4	26,7
<b>Rectal cancer (%)</b>		36,8	
<b>Colon cancer (%)</b>		63,2	
<b>Mutation</b>	<i>TP53</i>	77,0	
	<i>ras</i>	36,7	
<b>Family history of colorectal cancer</b>		11,7	3,8
<b>Smoking</b>	<b>Non smokers (%)</b>	53,2	55,9
	<b>Ex-smokers (%)</b>	31,6	26,1
	<b>Smokers (%)</b>	15,2	18,0
<b>Alcohol drinker (&gt; 10g/week)</b>	<b>never</b>	34,6	43,3
	<b>&lt;40 years</b>	31,7	30,5
	<b>&gt;40 years</b>	35,3	25,0
<b>Caloric intake</b>	<b>&lt;1688</b>	29,2	37,9
	<b>1688-2246</b>	34,4	31,9
	<b>≥2246</b>	36,3	30,2
<b>Use of drugs (weekly during &gt;6 months)</b>	<b>Acetylsalicylic acid</b>	16,1	17,0
	<b>Other NSAIDs</b>	5,8	14,5
	<b>Paracetamol</b>	10,5	17,4
	<b>Antiulcer</b>	16,9	13,6
	<b>Laxatives</b>	17,2	14,5
<b>BMI (at diagnosis)</b>	<b>&gt;25</b>	44,5	38,5
	<b>25-29</b>	41,5	39,4
	<b>≥30</b>	14,0	22,1
<b>BMI (10 years before diagnosis)</b>	<b>&gt;25</b>	36,2	37,5
	<b>25-29</b>	43,8	43,0
	<b>≥30</b>	20,0	19,5

Table 6. Characteristics of cases and control subjects: German population

		<b>Cases</b> (n=659) %	<b>Controls</b> (n=607) %
<b>Males/Females (%)</b>		58,4/41,6	56.8/43.2
<b>Age</b>	<b>33-54</b>	9,9	11,2
	<b>55-64</b>	25,3	28,8
	<b>65-74</b>	36,4	37,9
	<b>75-94</b>	28,4	22,1
<b>Rectal cancer (%)</b>		37,5	
<b>Colon cancer (%)</b>		62,2	
<b>First-degree family history of colorectal cancer</b>	<b>No</b>	79,2	79,9
	<b>Yes</b>	12,9	11,2
	<b>Unknown</b>	7,6	8,9
<b>UICC stage at diagnosis</b>	<b>I</b>	21,7	
	<b>II</b>	32,5	
	<b>III</b>	28,1	
	<b>IV</b>	17,3	

Table 7a. Logistic regression analysis for the selected SNPs in the miRNA target sequences. Only the co-dominant model is shown.

Genes	SNP Ref.	CASES <sup>A</sup>			CONTROLS <sup>A</sup>			OR HOMOZYGOTES			OR HETEROZYGOTES		
		WT/WT	WT/MUT	MUT/MUT	WT/WT	WT/MUT	MUT/MUT	OR <sup>B</sup>	LOW 95%	UP 95%	OR <sup>B</sup>	LOW 95%	UP 95%
<i>IL1A</i>	rs3783553	293	173	19	333	209	16	0,761	0,378	1,534	1,087	0,836	1,411
<i>IL12B</i>	rs1368439	446	183	21	417	174	14	1,402	0,692	2,840	0,965	0,749	1,243
<i>NOD2</i>	rs3135500	234	294	114	228	282	85	1,265	0,898	1,783	0,977	0,759	1,256
<i>PLA2G2A</i>	rs11677	521	118	10	468	114	8	1,201	0,451	3,196	0,917	0,684	1,230
<i>IL18</i>	rs360727	651	21	2	581	29	1	1,880	0,160	22,220	0,740	0,410	1,330
<i>CD86</i>	rs17281995	459	156	24	464	117	9	<b>2,740</b>	<b>1,243</b>	<b>6,041</b>	<b>1,384</b>	<b>1,047</b>	<b>1,829</b>
<i>PTGER4</i>	rs16870224	505	127	10	464	122	4	2,624	0,784	8,781	0,978	0,735	1,303
<i>IL16</i>	rs1131445	298	276	63	259	257	58	0,951	0,636	1,423	0,957	0,749	1,221
<i>ALOX15</i>	rs916055	270	307	67	231	308	57	1,050	0,701	1,574	0,832	0,653	1,061
<i>RAF1</i>	rs1051208	474	158	12	426	165	13	0,720	0,318	1,628	0,878	0,676	1,140
<i>CDKN2A</i>	rs11515	462	160	14	454	125	9	1,442	0,608	3,420	1,313	0,998	1,729
<i>CD4</i>	rs3213428	453	159	14	428	130	7	1,561	0,617	3,950	1,072	0,816	1,410
<i>INSR</i>	rs1051690	428	187	33	392	197	15	<b>1,942</b>	<b>1,031</b>	<b>3,660</b>	0,874	0,681	1,121
<i>INSR</i>	rs2860171	615	32	1	581	27	1	0,995	0,061	16,269	1,117	0,652	1,913
<i>GHR</i>	rs2973016	617	9	0	586	11	0	0,782	0,314	1,943	0,923	0,731	1,164

<sup>A</sup> Numbers may not add up to 100% of subjects due to genotyping failure. All samples that did not give a reliable result in the first round of genotyping were resubmitted to up to three additional rounds of genotyping. Data points that were still not filled after this procedure were left blank.

<sup>B</sup> OR = odds ratio; CI = confidence interval. ORs were adjusted for sex and age. Only results for the co-dominant model are shown. Statistically significant results (p<0.05) are shown in bold.

Table 7b. Logistic regression analysis for the selected SNPs in the miRNA sequences. Only the co-dominant model is shown.

		CASES <sup>A</sup>			CONTROLS <sup>A</sup>			OR HOMOZYGOTES			OR HETEROZIGOTES			
Genes	SNP Ref.	WT/WT	WT/MUT	MUT/MUT	WT/WT	WT/	MUT	MUT/MUT	OR <sup>B</sup>	LOW 95%	UP 95%	OR <sup>B</sup>	LOW 95%	UP 95%
miR-608	rs4919510	448	190	19	397	194		17	0,930	0,468	1,840	0,846	0,660	1,085
miR-492	rs2289030	573	68	7	529	70		4	1,714	0,479	6,130	0,867	0,603	1,248
miR-196a-2	rs11614913	282	263	92	229	273		78	0,986	0,689	1,411	0,778	0,606	6,998
miR-146a	rs2910164	389	217	46	354	189		46	0,952	0,611	1,485	0,985	0,768	1,263
miR-633	rs17759989	637	19	1	574	26		0	44044,0	undetermined		0,600	0,320	1,125
miR-605	rs2043555	360	249	33	337	204		33	0,999	0,596	1,675	1,197	0,651	1,529
miR-499	rs3746444	348	171	36	298	168		10	2,810	1,355	5,824	0,863	0,658	1,130
miR-631	rs5745925	581	64	6	485	67		1	4,957	0,577	42,531	0,801	0,553	1,139

<sup>A</sup> Numbers may not add up to 100% of subjects due to genotyping failure. All samples that did not give a reliable result in the first round of genotyping were resubmitted to up to three additional rounds of genotyping. Data points that were still not filled after this procedure were left blank.

<sup>B</sup> OR = odds ratio; CI = confidence interval. ORs were adjusted for sex and age. Only results for the co-dominant model are shown. Statistically significant results ( $p < 0.05$ ) are shown in bold.

*Table8.* Logistic regression analysis for the SNPs in CD86 and INSR genes: Spain population. Only the co-dominant model is shown.

		CASES <sup>A</sup>			CONTROLS <sup>A</sup>			OR HOMOZYGOTES			OR HETEROZIGOTES		
Genes	SNP	WT/WT	WT/MUT	MUT/MUT	WT/WT	WT/ MUT	MUT/MUT	OR <sup>B</sup>	LOW 95%	UP 95%	OR <sup>B</sup>	LOW 95%	UP 95%
<b>CD86</b>	rs17281995	238	82	12	181	67	7	1,34	0,51	3,48	0,97	0,66	1,42
<b>INSR</b>	rs1051690	232	88	17	195	75	4	3,20	1,05	9,78	1,00	0,69	1,44

<sup>A</sup> Numbers may not add up to 100% of subjects due to genotyping failure. All samples that did not give a reliable result in the first round of genotyping were resubmitted to up to three additional rounds of genotyping. Data points that were still not filled after this procedure were left blank.

<sup>B</sup> OR = odds ratio; CI = confidence interval. ORs were adjusted for sex and age. Only results for the co-dominant model are shown. Statistically significant results ( $p < 0.05$ ) are shown in bold.

*Table9.* Logistic regression analysis for the SNPs in CD86 and INSR genes: Italian population. Only the co-dominant model is shown.

		CASES <sup>A</sup>			CONTROLS <sup>A</sup>			OR HOMOZYGOTES			OR HETEROZIGOTES		
Genes	SNP	WT/WT	WT/MUT	MUT/MUT	WT/WT	WT/ MUT	MUT MUT/MUT	OR <sup>B</sup>	LOW 95%	UP 95%	OR <sup>B</sup>	LOW 95%	UP 95%
<b>CD86</b>	rs17281995	157	64	3	334	107	8	0,30	0,04	2,32	1,18	0,68	2,02
<b>INSR</b>	rs1051690	171	44	8	330	107	8	1,74	0,29	10,48	1,09	0,54	2,24

<sup>A</sup> Numbers may not add up to 100% of subjects due to genotyping failure. All samples that did not give a reliable result in the first round of genotyping were resubmitted to up to three additional rounds of genotyping. Data points that were still not filled after this procedure were left blank.

<sup>B</sup> OR = odds ratio; CI = confidence interval. ORs were adjusted for sex and age. Only results for the co-dominant model are shown. Statistically significant results ( $p < 0.05$ ) are shown in bold.

*Table 10.* Logistic regression analysis for the SNPs in CD86 and INSR genes: German population. Only the co-dominant model is shown.

		CASES <sup>A</sup>			CONTROLS <sup>A</sup>			OR HOMOZYGOTES			OR HETEROZIGOTES		
Genes	SNP	WT/WT	WT/MUT	MUT/MUT	WT/WT	WT/ MUT	MUT/MUT	OR <sup>B</sup>	LOW 95%	UP 95%	OR <sup>B</sup>	LOW 95%	UP 95%
<b>CD86</b>	rs17281995	853	329	29	921	310	31	1,01	0,60	1,69	1,15	0,95	1,37
<b>INSR</b>	rs1051690	845	349	36	887	343	45	0,84	0,54	1,31	1,10	0,90	1,30

<sup>A</sup> Numbers may not add up to 100% of subjects due to genotyping failure. All samples that did not give a reliable result in the first round of genotyping were resubmitted to up to three additional rounds of genotyping. Data points that were still not filled after this procedure were left blank.

<sup>B</sup> OR = odds ratio; CI = confidence interval. ORs were adjusted for sex and age. Only results for the co-dominant model are shown. Statistically significant results ( $p < 0.05$ ) are shown in bold.



*Table 11.* General logistic regression analysis for the SNPs in CD86 and INSR genes: Czech Republic, Spain, Italian, and German populations. Only the co-dominant model is shown.

		CASES <sup>A</sup>			CONTROLS <sup>A</sup>			OR HOMOZYGOTES			OR HETEROZIGOTES		
Genes	SNP	WT/WT	WT/MUT	MUT/MUT	WT/WT	WT/ MUT	MUT/MUT	OR <sup>B</sup>	LOW 95%	UP 95%	OR <sup>B</sup>	LOW 95%	UP 95%
<b>CD86</b>	rs17281995	1684	620	68	1870	598	54	<b>1,40</b>	<b>1,02</b>	<b>2,01</b>	<b>1,15</b>	<b>1,01</b>	<b>1,31</b>
<b>INSR</b>	rs1051690	1651	661	92	1776	718	70	<b>1,41</b>	<b>1,01</b>	<b>1,94</b>	0,99	0,87	1,12

<sup>A</sup> Numbers may not add up to 100% of subjects due to genotyping failure. All samples that did not give a reliable result in the first round of genotyping were resubmitted to up to three additional rounds of genotyping. Data points that were still not filled after this procedure were left blank.

<sup>B</sup> OR = odds ratio; CI = confidence interval. ORs were adjusted for sex and age. Only results for the co-dominant model are shown. Statistically significant results ( $p < 0.05$ ) are shown in bold.

## BIBLIOGRAPHY

1. Abelson, J.F., Kwan, K.Y., O'Roak, B.J., Baek, D.Y., Stillman, A.A., Morgan, T.M., Mathews, C.A., Pauls, D.L., Rasin, M.R., Gunel, M., Davis, N.R., Ercan-Sencicek, A.G., Guez, D.H., Spertus, J.A., Leckman, J.F., Dure, L.S., Kurlan, R., Singer, H.S., Gilbert, D.L., Farhi, A., Louvi, A., Lifton, R.P., Sestan, N., and State MW (2005). Sequence variants in *SLITRK1* are associated with Tourette's syndrome. *Science*. 310, 317-320.
2. Akao, Y., Nakagawa, Y., and Naoe, T. (2006). *let-7* microRNA functions as a potential growth suppressor in human colon cancer cells. *Biol. Pharm. Bull.* 29, 903-906.
3. Ambros, V., Bartel, B., Bartel, D.P., Burge, C.B., Carrington, J.C., Chen, X., Dreyfuss, G., Eddy, S.R., Griffiths-Jones, S., Marshall, M., Matzke, M., Ruvkun, G., and Tuschl, T. (2003a). A uniform system for microRNA annotation. *RNA*. 9, 277-279.
4. Ambros, V., Lee, R.C., Lavanway, A., Williams, P.T., and Jewell, D. (2003b). MicroRNAs and other tiny endogenous RNAs in *C. elegans*. *Curr. Biol.* 13, 807-818.
5. Ambros, V., Lee, R.C., Lavanway, A., Williams, P.T., and Jewell, D. (2003c). MicroRNAs and other tiny endogenous RNAs in *C. elegans*. *Curr. Biol.* 13, 807-818.
6. Amemori, S., Ootani, A., Aoki, S., Fujise, T., Shimoda, R., Kakimoto, T., Shiraishi, R., Sakata, Y., Tsunada, S., Iwakiri, R., and Fujimoto, K. (2007). Adipocytes and preadipocytes promote the proliferation of colon cancer cells in vitro. *Am. J. Physiol Gastrointest. Liver Physiol.* 292, G923-G929.
7. Ames, B.N. (2001). DNA damage from micronutrient deficiencies is likely to be a major cause of cancer. *Mutat. Res.* 475, 7-20.
8. Aravin, A.A., Hannon, G.J., and Brennecke, J. (2007). The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. *Science*. 318, 761-764.
9. Aravin, A.A., Naumova, N.M., Tulin, A.V., Vagin, V.V., Rozovsky, Y.M., and Gvozdev, V.A. (2001). Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the *D. melanogaster* germline. *Curr. Biol.* 11, 1017-1027.
10. Bartel, D.P. (2004b). MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116, 281-297.
11. Bartel, D.P. (2004a). MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116, 281-297.

12. Baskerville,S. and Bartel,D.P. (2005). Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. *RNA*. 11, 241-247.
13. Basyuk,E., Suavet,F., Doglio,A., Bordonne,R., and Bertrand,E. (2003). Human let-7 stem-loop precursors harbor features of RNase III cleavage products. *Nucleic Acids Res*. 31, 6593-6597.
14. Berezikov,E., Chung,W.J., Willis,J., Cuppen,E., and Lai,E.C. (2007). Mammalian mirtron genes. *Mol. Cell*. 28, 328-336.
15. Bernstein,E., Caudy,A.A., Hammond,S.M., and Hannon,G.J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature*. 409, 363-366.
16. Bhatia,S., Edidin,M., Almo,S.C., and Nathenson,S.G. (2006). B7-1 and B7-2: similar costimulatory ligands with different biochemical, oligomeric and signaling properties. *Immunol. Lett*. 104, 70-75.
17. Bingham,S.A., Pignatelli,B., Pollock,J.R., Ellul,A., Malaveille,C., Gross,G., Runswick,S., Cummings,J.H., and O'Neill,I.K. (1996). Does increased endogenous formation of N-nitroso compounds in the human colon explain the association between red meat and colon cancer? *Carcinogenesis*. 17, 515-523.
18. Borchert,G.M., Lanier,W., and Davidson,B.L. (2006). RNA polymerase III transcribes human microRNAs. *Nat. Struct. Mol. Biol*. 13, 1097-1101.
19. Boyle,P. and Langman,J.S. (2000). ABC of colorectal cancer: Epidemiology. *BMJ* 321, 805-808.
20. Brennecke,J., Stark,A., Russell,R.B., and Cohen,S.M. (2005). Principles of microRNA-target recognition. *PLoS. Biol*. 3, e85.
21. Calin,G.A., Dumitru,C.D., Shimizu,M., Bichi,R., Zupo,S., Noch,E., Aldler,H., Rattan,S., Keating,M., Rai,K., Rassenti,L., Kipps,T., Negrini,M., Bullrich,F., and Croce,C.M. (2002). Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proc. Natl. Acad. Sci. U. S. A* 99, 15524-15529.
22. Calin,G.A., Ferracin,M., Cimmino,A., Di,L.G., Shimizu,M., Wojcik,S.E., Iorio,M.V., Visone,R., Sever,N.I., Fabbri,M., Iuliano,R., Palumbo,T., Pichiorri,F., Roldo,C., Garzon,R., Sevignani,C., Rassenti,L., Alder,H., Volinia,S., Liu,C.G., Kipps,T.J., Negrini,M., and Croce,C.M. (2005). A MicroRNA signature associated with prognosis and progression in chronic lymphocytic leukemia. *N. Engl. J. Med*. 353, 1793-1801.
23. Carmell,M.A., Xuan,Z., Zhang,M.Q., and Hannon,G.J. (2002). The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. *Genes Dev*. 16, 2733-2742.

24. Carthew,R.W. (2006). Gene regulation by microRNAs. *Curr. Opin. Genet. Dev.* 16, 203-208.
25. Chan,J.A., Krichevsky,A.M., and Kosik,K.S. (2005). MicroRNA-21 is an antiapoptotic factor in human glioblastoma cells. *Cancer Res.* 65, 6029-6033.
26. Cimmino,A., Calin,G.A., Fabbri,M., Iorio,M.V., Ferracin,M., Shimizu,M., Wojcik,S.E., Aqeilan,R.I., Zupo,S., Dono,M., Rassenti,L., Alder,H., Volinia,S., Liu,C.G., Kipps,T.J., Negrini,M., and Croce,C.M. (2005). miR-15 and miR-16 induce apoptosis by targeting BCL2. *Proc. Natl. Acad. Sci. U. S. A.* 102, 13944-13949.
27. Clop,A., Marcq,F., Takeda,H., Pirottin,D., Tordoir,X., Bibe,B., Bouix,J., Caiment,F., Elsen,J.M., Eychenne,F., Larzul,C., Laville,E., Meish,F., Milenkovic,D., Tobin,J., Charlier,C., and Georges,M. (2006). A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. *Nat. Genet.* 38, 813-818.
28. Cross,A.J. and Sinha,R. (2004). Meat-related mutagens/carcinogens in the etiology of colorectal cancer. *Environ. Mol. Mutagen.* 44, 44-55.
29. Cullen,B.R. (2004). Transcription and processing of human microRNA precursors. *Mol. Cell* 16, 861-865.
30. Cummins,J.M., He,Y., Leary,R.J., Pagliarini,R., Diaz,L.A., Jr., Sjoblom,T., Barad,O., Bentwich,Z., Szafranska,A.E., Labourier,E., Raymond,C.K., Roberts,B.S., Juhl,H., Kinzler,K.W., Vogelstein,B., and Velculescu,V.E. (2006). The colorectal microRNAome. *Proc. Natl. Acad. Sci. U. S. A* 103, 3687-3692.
31. Duan,R., Pak,C., and Jin,P. (2007). Single nucleotide polymorphism associated with mature miR-125a alters the processing of pri-miRNA. *Hum. Mol. Genet.* 16, 1124-1131.
32. Dugas,D.V. and Bartel,B. (2004). MicroRNA regulation of gene expression in plants. *Curr. Opin. Plant Biol.* 7, 512-520.
33. Eis,P.S., Tam,W., Sun,L., Chadburn,A., Li,Z., Gomez,M.F., Lund,E., and Dahlberg,J.E. (2005). Accumulation of miR-155 and BIC RNA in human B cell lymphomas. *Proc. Natl. Acad. Sci. U. S. A.* 102, 3627-3632.
34. Esquela-Kerscher,A. and Slack,F.J. (2006). Oncomirs - microRNAs with a role in cancer. *Nat. Rev. Cancer.* 6, 259-269.
35. Esteller,M. and Herman,J.G. (2002). Cancer as an epigenetic disease: DNA methylation and chromatin alterations in human tumours. *J. Pathol.* 196, 1-7.
36. Fantuzzi,G. (2005). Adipose tissue, adipokines, and inflammation. *J. Allergy Clin. Immunol.* 115, 911-919.

37. Filipowicz,W., Jaskiewicz,L., Kolb,F.A., and Pillai,R.S. (2005). Post-transcriptional gene silencing by siRNAs and miRNAs. *Curr. Opin. Struct. Biol.* 15, 331-341.
38. Fire,A., Xu,S., Montgomery,M.K., Kostas,S.A., Driver,S.E., and Mello,C.C. (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature.* %19;391, 806-811.
39. Flood,D.M., Weiss,N.S., Cook,L.S., Emerson,J.C., Schwartz,S.M., and Potter,J.D. (2000). Colorectal cancer incidence in Asian migrants to the United States and their descendants. *Cancer Causes Control* 11, 403-411.
40. Foti,D., Iuliano,R., Chiefari,E., and Brunetti,A. (2003). A nucleoprotein complex containing Sp1, C/EBP beta, and HMGI-Y controls human insulin receptor gene transcription. *Mol. Cell Biol.* 23, 2720-2732.
41. Fukuda,T., Yamagata,K., Fujiyama,S., Matsumoto,T., Koshida,I., Yoshimura,K., Mihara,M., Naitou,M., Endoh,H., Nakamura,T., Akimoto,C., Yamamoto,Y., Katagiri,T., Foulds,C., Takezawa,S., Kitagawa,H., Takeyama,K., O'Malley,B.W., and Kato,S. (2007). DEAD-box RNA helicase subunits of the Drosha complex are required for processing of rRNA and a subset of microRNAs. *Nat. Cell Biol.* 9, 604-611.
42. Gafa,R. and Lanza,G. (1998). [Expression of protein p53 in the adenoma-colorectal carcinoma sequence]. *Pathologica.* 90, 351-356.
43. Garland,C.F. and Garland,F.C. (2006). Do sunlight and vitamin D reduce the likelihood of colon cancer? *Int. J. Epidemiol.* 35, 217-220.
44. Gauwerky,C.E., Huebner,K., Isobe,M., Nowell,P.C., and Croce,C.M. (1989). Activation of MYC in a masked t(8;17) translocation results in an aggressive B-cell leukemia. *Proc. Natl. Acad. Sci. U. S. A.* 86, 8867-8871.
45. Giovannucci,E. (2001). Insulin, insulin-like growth factors and colon cancer: a review of the evidence. *J. Nutr.* 131, 3109S-3120S.
46. Giovannucci,E., Stampfer,M.J., Colditz,G.A., Rimm,E.B., Trichopoulos,D., Rosner,B.A., Speizer,F.E., and Willett,W.C. (1993). Folate, methionine, and alcohol intake and risk of colorectal adenoma. *J. Natl. Cancer Inst.* 85, 875-884.
47. Gregory,R.I., Chendrimada,T.P., Cooch,N., and Shiekhattar,R. (2005). Human RISC couples microRNA biogenesis and posttranscriptional gene silencing. *Cell.* 123, 631-640.
48. Griffiths-Jones,S., Grocock,R.J., van,D.S., Bateman,A., and Enright,A.J. (2006). miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.* 34, D140-D144.

49. Guil,S. and Caceres,J.F. (2007). The multifunctional RNA-binding protein hnRNP A1 is required for processing of miR-18a. *Nat. Struct. Mol. Biol.* **14**, 591-596.
50. Gunter,M.J. and Leitzmann,M.F. (2006). Obesity and colorectal cancer: epidemiology, mechanisms and candidate genes. *J. Nutr. Biochem.* **17**, 145-156.
51. Gunter,M.J., Stolzenberg-Solomon,R., Cross,A.J., Leitzmann,M.F., Weinstein,S., Wood,R.J., Virtamo,J., Taylor,P.R., Albanes,D., and Sinha,R. (2006). A prospective study of serum C-reactive protein and colorectal cancer risk in men. *Cancer Res.* **66**, 2483-2487.
52. Hall,M.N., Campos,H., Li,H., Sesso,H.D., Stampfer,M.J., Willett,W.C., and Ma,J. (2007). Blood levels of long-chain polyunsaturated fatty acids, aspirin, and the risk of colorectal cancer. *Cancer Epidemiol. Biomarkers Prev.* **16**, 314-321.
53. Han,J., Lee,Y., Yeom,K.H., Kim,Y.K., Jin,H., and Kim,V.N. (2004). The Drosha-DGCR8 complex in primary microRNA processing. *Genes Dev.* **18**, 3016-3027.
54. Han,J., Lee,Y., Yeom,K.H., Nam,J.W., Heo,I., Rhee,J.K., Sohn,S.Y., Cho,Y., Zhang,B.T., and Kim,V.N. (2006). Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. *Cell* **125**, 887-901.
55. Haruta,T., Imamura,T., Iwanishi,M., Egawa,K., Goji,K., and Kobayashi,M. (1995). Amplification and analysis of promoter region of insulin receptor gene in a patient with leprechaunism associated with severe insulin resistance. *Metabolism.* **44**, 430-437.
56. He,L., Thomson,J.M., Hemann,M.T., Hernando-Monge,E., Mu,D., Goodson,S., Powers,S., Cordon-Cardo,C., Lowe,S.W., Hannon,G.J., and Hammond,S.M. (2005). A microRNA polycistron as a potential human oncogene. *Nature.* **435**, 828-833.
57. Humphreys,D.T., Westman,B.J., Martin,D.I., and Preiss,T. (2005). MicroRNAs control translation initiation by inhibiting eukaryotic initiation factor 4E/cap and poly(A) tail function. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 16961-16966.
58. Iorio,M.V., Ferracin,M., Liu,C.G., Veronese,A., Spizzo,R., Sabbioni,S., Magri,E., Pedriali,M., Fabbri,M., Campiglio,M., Menard,S., Palazzo,J.P., Rosenberg,A., Musiani,P., Volinia,S., Nenci,I., Calin,G.A., Querzoli,P., Negrini,M., and Croce,C.M. (2005). MicroRNA gene expression deregulation in human breast cancer. *Cancer Res.* **65**, 7065-7070.
59. Iwai,N. and Naraba,H. (2005). Polymorphisms in human pre-miRNAs. *Biochem. Biophys. Res. Commun.* **331**, 1439-1444.

60. John,B., Enright,A.J., Aravin,A., Tuschl,T., Sander,C., and Marks,D.S. (2004). Human MicroRNA targets. *PLoS. Biol.* 2, e363.
61. John,B.J., Irukulla,S., Abulafi,A.M., Kumar,D., and Mendall,M.A. (2006). Systematic review: adipose tissue, obesity and gastrointestinal diseases. *Aliment. Pharmacol. Ther.* 23, 1511-1523.
62. Johnson,I.T. (2001). Mechanisms and anticarcinogenic effects of diet-related apoptosis in the intestinal mucosa. *Nutr. Res. Rev.* 14, 229-256.
63. Johnson,S.M., Grosshans,H., Shingara,J., Byrom,M., Jarvis,R., Cheng,A., Labourier,E., Reinert,K.L., Brown,D., and Slack,F.J. (2005). RAS is regulated by the let-7 microRNA family. *Cell.* 120, 635-647.
64. Kadowaki,T., Kadowaki,H., Rechler,M.M., Serrano-Rios,M., Roth,J., Gorden,P., and Taylor,S.I. (1990). Five mutant alleles of the insulin receptor gene in patients with genetic forms of insulin resistance. *J. Clin. Invest.* 86, 254-264.
65. Keku,T.O., Lund,P.K., Galanko,J., Simmons,J.G., Woosley,J.T., and Sandler,R.S. (2005). Insulin resistance, apoptosis, and colorectal adenoma risk. *Cancer Epidemiol. Biomarkers Prev.* 14, 2076-2081.
66. Kelly,D. and Conway,S. (2005). Bacterial modulation of mucosal innate immunity. *Mol. Immunol.* 42, 895-901.
67. Kim,V.N. (2006). Small RNAs just got bigger: Piwi-interacting RNAs (piRNAs) in mammalian testes. *Genes Dev.* 20, 1993-1997.
68. Kiriakidou,M., Nelson,P.T., Kouranov,A., Fitziev,P., Bouyioukos,C., Mourelatos,Z., and Hatzigeorgiou,A. (2004). A combined computational-experimental approach predicts human microRNA targets. *Genes Dev.* 18, 1165-1178.
69. Krek,A., Grun,D., Poy,M.N., Wolf,R., Rosenberg,L., Epstein,E.J., MacMenamin,P., da,P., I, Gunsalus,K.C., Stoffel,M., and Rajewsky,N. (2005). Combinatorial microRNA target predictions. *Nat. Genet.* 37, 495-500.
70. Krutzfeldt,J., Poy,M.N., and Stoffel,M. (2006). Strategies to determine the biological function of microRNAs. *Nat. Genet.* 38 *Suppl*:S14-9., S14-S19.
71. Kutay,H., Bai,S., Datta,J., Motiwala,T., Pogribny,I., Frankel,W., Jacob,S.T., and Ghoshal,K. (2006). Downregulation of miR-122 in the rodent and human hepatocellular carcinomas. *J. Cell Biochem.* 99, 671-678.
72. Lagos-Quintana,M., Rauhut,R., Meyer,J., Borkhardt,A., and Tuschl,T. (2003). New microRNAs from mouse and human. *RNA.* 9, 175-179.

73. Lai,E.C. (2002). Micro RNAs are complementary to 3' UTR sequence motifs that mediate negative post-transcriptional regulation. *Nat. Genet.* 30, 363-364.
74. Lamlum,H., Papadopoulou,A., Ilyas,M., Rowan,A., Gillet,C., Hanby,A., Talbot,I., Bodmer,W., and Tomlinson,I. (2000). APC mutations are sufficient for the growth of early colorectal adenomas. *Proc. Natl. Acad. Sci. U. S. A.* 97, 2225-2228.
75. Law,M.E., Templeton,K.L., Kitange,G., Smith,J., Misra,A., Feuerstein,B.G., and Jenkins,R.B. (2005). Molecular cytogenetic analysis of chromosomes 1 and 19 in glioma cell lines. *Cancer Genet. Cytogenet.* 160, 1-14.
76. Lee,R.C. and Ambros,V. (2001). An extensive class of small RNAs in *Caenorhabditis elegans*. *Science* 294, 862-864.
77. Lee,R.C., Feinbaum,R.L., and Ambros,V. (1993b). The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell.* 75, 843-854.
78. Lee,R.C., Feinbaum,R.L., and Ambros,V. (1993a). The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* 75, 843-854.
79. Lee,Y., Kim,M., Han,J., Yeom,K.H., Lee,S., Baek,S.H., and Kim,V.N. (2004). MicroRNA genes are transcribed by RNA polymerase II. *EMBO J.* 23, 4051-4060.
80. Lewis,B.P., Shih,I.H., Jones-Rhoades,M.W., Bartel,D.P., and Burge,C.B. (2003). Prediction of mammalian microRNA targets. *Cell.* 115, 787-798.
81. Lim,L.P., Glasner,M.E., Yekta,S., Burge,C.B., and Bartel,D.P. (2003). Vertebrate microRNA genes. *Science.* 299, 1540.
82. Lin,H.J., Probst-Hensch,N.M., Louie,A.D., Kau,I.H., Witte,J.S., Ingles,S.A., Frankl,H.D., Lee,E.R., and Haile,R.W. (1998). Glutathione transferase null genotype, broccoli, and lower prevalence of colorectal adenomas. *Cancer Epidemiol. Biomarkers Prev.* 7, 647-652.
83. Lindbo,J.A., Silva-Rosales,L., Proebsting,W.M., and Dougherty,W.G. (1993). Induction of a Highly Specific Antiviral State in Transgenic Plants: Implications for Regulation of Gene Expression and Virus Resistance. *Plant Cell.* 5, 1749-1759.
84. Lingel,A., Simon,B., Izaurralde,E., and Sattler,M. (2004). Nucleic acid 3'-end recognition by the Argonaute2 PAZ domain. *Nat. Struct. Mol. Biol.* 11, 576-577.
85. Linsley,P.S., Brady,W., Grosmaire,L., Aruffo,A., Damle,N.K., and Ledbetter,J.A. (1991). Binding of the B cell activation antigen B7 to CD28



- costimulates T cell proliferation and interleukin 2 mRNA accumulation. *J. Exp. Med.* 173, 721-730.
86. Lipkin,M., Reddy,B., Newmark,H., and Lamprecht,S.A. (1999). Dietary factors in human colorectal cancer. *Annu. Rev. Nutr.* 19:545-86., 545-586.
  87. Liu,J., Carmell,M.A., Rivas,F.V., Marsden,C.G., Thomson,J.M., Song,J.J., Hammond,S.M., Joshua-Tor,L., and Hannon,G.J. (2004). Argonaute2 is the catalytic engine of mammalian RNAi. *Science.* 305, 1437-1441.
  88. Lund,E.K., Wharf,S.G., Fairweather-Tait,S.J., and Johnson,I.T. (1999). Oral ferrous sulfate supplements increase the free radical-generating capacity of feces from healthy volunteers. *Am. J. Clin. Nutr.* 69, 250-255.
  89. MacInnis,R.J., English,D.R., Hopper,J.L., Gertig,D.M., Haydon,A.M., and Giles,G.G. (2006). Body size and composition and colon cancer risk in women. *Int. J. Cancer.* 118, 1496-1500.
  90. Maniataki,E. and Mourelatos,Z. (2005). A human, ATP-independent, RISC assembly machine fueled by pre-miRNA. *Genes Dev.* 19, 2979-2990.
  91. Martin,M.M., Buckenberger,J.A., Jiang,J., Malana,G.E., Nuovo,G.J., Chotani,M., Feldman,D.S., Schmittgen,T.D., and Elton,T.S. (2007). The human angiotensin II type 1 receptor +1166 A/C polymorphism attenuates microrna-155 binding. *J. Biol. Chem.* 282, 24262-24269.
  92. Martinez,J. and Tuschl,T. (2004). RISC is a 5' phosphomonoester-producing RNA endonuclease. *Genes Dev.* 18, 975-980.
  93. Michael,M.Z., O' Connor,S.M., van Holst Pellekaan,N.G., Young,G.P., and James,R.J. (2003). Reduced accumulation of specific microRNAs in colorectal neoplasia. *Mol. Cancer Res.* 1, 882-891.
  94. Moore,L.L., Bradlee,M.L., Singer,M.R., Splansky,G.L., Proctor,M.H., Ellison,R.C., and Kreger,B.E. (2004). BMI and waist circumference as predictors of lifetime colon cancer risk in Framingham Study adults. *Int. J. Obes. Relat Metab Disord.* 28, 559-567.
  95. Mourelatos,Z., Dostie,J., Paushkin,S., Sharma,A., Charroux,B., Abel,L., Rappsilber,J., Mann,M., and Dreyfuss,G. (2002). miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. *Genes Dev.* 16, 720-728.
  96. Munkholm,P. (2003). Review article: the incidence and prevalence of colorectal cancer in inflammatory bowel disease. *Aliment. Pharmacol. Ther.* 18 Suppl 2:1-5., 1-5.

97. Nelson,P.T., Hatzigeorgiou,A.G., and Mourelatos,Z. (2004). miRNP:mRNA association in polyribosomes in a human neuronal cell line. *RNA*. 10, 387-394.
98. Norat,T., Bingham,S., Ferrari,P., Slimani,N., Jenab,M., Mazuir,M., Overvad,K., Olsen,A., Tjonneland,A., Clavel,F., Boutron-Ruault,M.C., Kesse,E., Boeing,H., Bergmann,M.M., Nieters,A., Linseisen,J., Trichopoulou,A., Trichopoulos,D., Tountas,Y., Berrino,F., Palli,D., Panico,S., Tumino,R., Vineis,P., Bueno-de-Mesquita,H.B., Peeters,P.H., Engeset,D., Lund,E., Skeie,G., Ardanaz,E., Gonzalez,C., Navarro,C., Quiros,J.R., Sanchez,M.J., Berglund,G., Mattisson,I., Hallmans,G., Palmqvist,R., Day,N.E., Khaw,K.T., Key,T.J., San,J.M., Hemon,B., Saracci,R., Kaaks,R., and Riboli,E. (2005). Meat, fish, and colorectal cancer risk: the European Prospective Investigation into cancer and nutrition. *J. Natl. Cancer Inst.* 97, 906-916.
99. Nottrott,S., Simard,M.J., and Richter,J.D. (2006). Human let-7a miRNA blocks protein production on actively translating polyribosomes. *Nat. Struct. Mol. Biol.* 13, 1108-1114.
100. Novina,C.D. and Sharp,P.A. (2004). The RNAi revolution. *Nature* 430, 161-164.
101. O'Donnell,K.A., Wentzel,E.A., Zeller,K.I., Dang,C.V., and Mendell,J.T. (2005). c-Myc-regulated microRNAs modulate E2F1 expression. *Nature*. 435, 839-843.
102. Okabayashi,Y., Maddux,B.A., McDonald,A.R., Logsdon,C.D., Williams,J.A., and Goldfine,I.D. (1989b). Mechanisms of insulin-induced insulin-receptor downregulation. Decrease of receptor biosynthesis and mRNA levels. *Diabetes* 38, 182-187.
103. Okabayashi,Y., Maddux,B.A., McDonald,A.R., Logsdon,C.D., Williams,J.A., and Goldfine,I.D. (1989a). Mechanisms of insulin-induced insulin-receptor downregulation. Decrease of receptor biosynthesis and mRNA levels. *Diabetes* 38, 182-187.
104. Okamura,K., Hagen,J.W., Duan,H., Tyler,D.M., and Lai,E.C. (2007). The mirtron pathway generates microRNA-class regulatory RNAs in *Drosophila*. *Cell*. 130, 89-100.
105. Orabona,C., Grohmann,U., Belladonna,M.L., Fallarino,F., Vacca,C., Bianchi,R., Bozza,S., Volpi,C., Salomon,B.L., Fioretti,M.C., Romani,L., and Puccetti,P. (2004). CD28 induces immunostimulatory signals in dendritic cells via CD80 and CD86. *Nat. Immunol.* 5, 1134-1142.
106. Otani,T., Iwasaki,M., Sasazuki,S., Inoue,M., and Tsugane,S. (2006). Plasma C-reactive protein and risk of colorectal cancer in a nested case-control study: Japan Public Health Center-based prospective study. *Cancer Epidemiol. Biomarkers Prev.* 15, 690-695.

107. Pallante,P., Visone,R., Ferracin,M., Ferraro,A., Berlingieri,M.T., Troncone,G., Chiappetta,G., Liu,C.G., Santoro,M., Negrini,M., Croce,C.M., and Fusco,A. (2006). MicroRNA deregulation in human thyroid papillary carcinomas. *Endocr. Relat Cancer* 13, 497-508.
108. Parkin,D.M., Bray,F., Ferlay,J., and Pisani,P. (2005a). Global cancer statistics, 2002. *CA Cancer J. Clin.* 55, 74-108.
109. Parkin,D.M., Bray,F., Ferlay,J., and Pisani,P. (2005b). Global cancer statistics, 2002. *CA Cancer J. Clin.* 55, 74-108.
110. Pasquinelli,A.E., Reinhart,B.J., Slack,F., Martindale,M.Q., Kuroda,M.I., Maller,B., Hayward,D.C., Ball,E.E., Degan,B., Muller,P., Spring,J., Srinivasan,A., Fishman,M., Finnerty,J., Corbo,J., Levine,M., Leahy,P., Davidson,E., and Ruvkun,G. (2000). Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature.* 408, 86-89.
111. Pechlivanis,S., Pardini,B., Bermejo,J.L., Wagner,K., Naccarati,A., Vodickova,L., Novotny,J., Hemminki,K., Vodicka,P., and Forsti,A. (2007). Insulin pathway related genes and risk of colorectal cancer: INSR promoter polymorphism shows a protective effect. *Endocr. Relat Cancer.* 14, 733-740.
112. Petersen,C.P., Bordeleau,M.E., Pelletier,J., and Sharp,P.A. (2006). Short RNAs repress translation after initiation in mammalian cells. *Mol. Cell.* 21, 533-542.
113. Petri,N., Tannergren,C., Holst,B., Mellon,F.A., Bao,Y., Plumb,G.W., Bacon,J., O'Leary,K.A., Kroon,P.A., Knutson,L., Forsell,P., Eriksson,T., Lennernas,H., and Williamson,G. (2003). Absorption/metabolism of sulforaphane and quercetin, and regulation of phase II enzymes, in human jejunum in vivo. *Drug Metab Dispos.* 31, 805-813.
114. Pillai,R.S., Bhattacharyya,S.N., Artus,C.G., Zoller,T., Cougot,N., Basyuk,E., Bertrand,E., and Filipowicz,W. (2005). Inhibition of translational initiation by Let-7 MicroRNA in human cells. *Science.* 309, 1573-1576.
115. Pirola,L., Johnston,A.M., and Van,O.E. (2004). Modulation of insulin action. *Diabetologia* 47, 170-184.
116. Pischon,T., Lahmann,P.H., Boeing,H., Friedenreich,C., Norat,T., Tjonneland,A., Halkjaer,J., Overvad,K., Clavel-Chapelon,F., Boutron-Ruault,M.C., Guernec,G., Bergmann,M.M., Linseisen,J., Becker,N., Trichopoulou,A., Trichopoulos,D., Sieri,S., Palli,D., Tumino,R., Vineis,P., Panico,S., Peeters,P.H., Bueno-de-Mesquita,H.B., Boshuizen,H.C., Van,G.B., Palmqvist,R., Berglund,G., Gonzalez,C.A., Dorronsoro,M., Barricarte,A., Navarro,C., Martinez,C., Quiros,J.R., Roddam,A., Allen,N., Bingham,S., Khaw,K.T., Ferrari,P., Kaaks,R., Slimani,N., and Riboli,E. (2006). Body size and risk of colon and rectal cancer in the European

Prospective Investigation Into Cancer and Nutrition (EPIC). *J. Natl. Cancer Inst.* 98, 920-931.

117. Provost,P., Dishart,D., Doucet,J., Frendewey,D., Samuelsson,B., and Radmark,O. (2002). Ribonuclease activity and RNA binding of recombinant human Dicer. *EMBO J.* 21, 5864-5874.
118. Puig,O. and Tjian,R. (2005). Transcriptional feedback control of insulin receptor by dFOXO/FOXO1. *Genes Dev.* 19, 2435-2446.
119. Rajewsky,N. (2006b). microRNA target predictions in animals. *Nat. Genet.* 38 *Suppl*:S8-13., S8-13.
120. Rajewsky,N. (2006a). L(ou)sy miRNA targets? *Nat. Struct. Mol. Biol.* 13, 754-755.
121. Rodriguez,A., Griffiths-Jones,S., Ashurst,J.L., and Bradley,A. (2004). Identification of mammalian microRNA host genes and transcription units. *Genome Res.* 14, 1902-1910.
122. Roldo,C., Missiaglia,E., Hagan,J.P., Falconi,M., Capelli,P., Bersani,S., Calin,G.A., Volinia,S., Liu,C.G., Scarpa,A., and Croce,C.M. (2006). MicroRNA expression abnormalities in pancreatic endocrine and acinar tumors are associated with distinctive pathologic features and clinical behavior. *J. Clin. Oncol.* 24, 4677-4684.
123. Rome,S., Meugnier,E., and Vidal,H. (2004). The ubiquitin-proteasome pathway is a new partner for the control of insulin signaling. *Curr. Opin. Clin. Nutr. Metab Care* 7, 249-254.
124. Rondinone,C.M. (2006). Adipocyte-derived hormones, cytokines, and mediators. *Endocrine.* 29, 81-90.
125. Ross,R., Leger,L., Morris,D., de,G.J., and Guardo,R. (1992). Quantification of adipose tissue by MRI: relationship with anthropometric variables. *J. Appl. Physiol.* 72, 787-795.
126. Ruby,J.G., Jan,C.H., and Bartel,D.P. (2007). Intronic microRNA precursors that bypass Drosha processing. *Nature.* 448, 83-86.
127. Rusinov,V., Baev,V., Minkov,I.N., and Tabler,M. (2005). MicroInspector: a web tool for detection of miRNA binding sites in an RNA sequence. *Nucleic Acids Res.* 33, W696-W700.
128. Saito,Y., Liang,G., Egger,G., Friedman,J.M., Chuang,J.C., Coetzee,G.A., and Jones,P.A. (2006). Specific activation of microRNA-127 with downregulation of the proto-oncogene BCL6 by chromatin-modifying drugs in human cancer cells. *Cancer Cell.* 9, 435-443.
129. Samanic,C., Chow,W.H., Gridley,G., Jarvholm,B., and Fraumeni,J.F., Jr. (2006). Relation of body mass index to cancer risk in 362,552 Swedish men. *Cancer Causes Control.* 17, 901-909.

130. Scott,G.K., Goga,A., Bhaumik,D., Berger,C.E., Sullivan,C.S., and Benz,C.C. (2007). Coordinate suppression of ERBB2 and ERBB3 by enforced expression of micro-RNA miR-125a or miR-125b. *J. Biol. Chem.* 282, 1479-1486.
131. Seggerson,K., Tang,L., and Moss,E.G. (2002). Two genetic circuits repress the *Caenorhabditis elegans* heterochronic gene *lin-28* after translation initiation. *Dev. Biol.* 243, 215-225.
132. Sethupathy,P., Borel,C., Gagnebin,M., Grant,G.R., Deutsch,S., Elton,T.S., Hatzigeorgiou,A.G., and Antonarakis,S.E. (2007). Human microRNA-155 on chromosome 21 differentially interacts with its polymorphic target in the *AGTR1* 3' untranslated region: a mechanism for functional single-nucleotide polymorphisms related to phenotypes. *Am. J. Hum. Genet.* 81, 405-413.
133. Siezen,C.L., Bueno-de-Mesquita,H.B., Peeters,P.H., Kram,N.R., van,D.M., and van Kranen,H.J. (2006). Polymorphisms in the genes involved in the arachidonic acid-pathway, fish consumption and the risk of colorectal cancer. *Int. J. Cancer.* 119, 297-303.
134. Smalheiser,N.R. (2003). EST analyses predict the existence of a population of chimeric microRNA precursor-mRNA transcripts expressed in normal human and mouse tissues. *Genome Biol.* 4, 403.
135. Smalheiser,N.R. and Torvik,V.I. (2005). Mammalian microRNAs derived from genomic repeats. *Trends Genet.* 21, 322-326.
136. Sullivan,C.S. and Ganem,D. (2005). MicroRNAs and viral infection. *Mol. Cell.* 20, 3-7.
137. Suvas,S., Singh,V., Sahdev,S., Vohra,H., and Agrewala,J.N. (2002). Distinct role of CD80 and CD86 in the regulation of the activation of B cell and B cell lymphoma. *J. Biol. Chem.* 277, 7766-7775.
138. Tam,W., Ben-Yehuda,D., and Hayward,W.S. (1997). *bic*, a novel gene activated by proviral insertions in avian leukosis virus-induced lymphomas, is likely to function through its noncoding RNA. *Mol. Cell Biol.* 17, 1490-1502.
139. Taniguchi,C.M., Emanuelli,B., and Kahn,C.R. (2006). Critical nodes in signalling pathways: insights into insulin action. *Nat. Rev. Mol. Cell Biol.* 7, 85-96.
140. Tanzer,A. and Stadler,P.F. (2004). Molecular evolution of a microRNA cluster. *J. Mol. Biol.* 339, 327-335.
141. Tivol,E.A., Borriello,F., Schweitzer,A.N., Lynch,W.P., Bluestone,J.A., and Sharpe,A.H. (1995). Loss of CTLA-4 leads to massive lymphoproliferation and fatal multiorgan tissue destruction, revealing a critical negative regulatory role of CTLA-4. *Immunity.* 3, 541-547.

142. Tomari,Y. and Zamore,P.D. (2005). Perspective: machines for RNAi. *Genes Dev.* 19, 517-529.
143. Tuynman,J.B., Peppelenbosch,M.P., and Richel,D.J. (2004). COX-2 inhibition as a tool to treat and prevent colorectal cancer. *Crit Rev. Oncol. Hematol.* 52, 81-101.
144. van,P.G., Verhoeven,D.T., Verhagen,H., and Goldbohm,R.A. (1999). Brassica vegetables and cancer prevention. *Epidemiology and mechanisms. Adv. Exp. Med. Biol.* 472:159-68., 159-168.
145. Vella,M.C., Choi,E.Y., Lin,S.Y., Reinert,K., and Slack,F.J. (2004). The *C. elegans* microRNA let-7 binds to imperfect let-7 complementary sites from the lin-41 3'UTR. *Genes Dev.* 18, 132-137.
146. Vogelstein,B., Fearon,E.R., Hamilton,S.R., Kern,S.E., Preisinger,A.C., Leppert,M., Nakamura,Y., White,R., Smits,A.M., and Bos,J.L. (1988). Genetic alterations during colorectal-tumor development. *N. Engl. J. Med.* 319, 525-532.
147. Vuckovic,S., Florin,T.H., Khalil,D., Zhang,M.F., Patel,K., Hamilton,I., and Hart,D.N. (2001). CD40 and CD86 upregulation with divergent CMRF44 expression on blood dendritic cells in inflammatory bowel diseases. *Am. J. Gastroenterol.* 96, 2946-2956.
148. Wei,E.K., Ma,J., Pollak,M.N., Rifai,N., Fuchs,C.S., Hankinson,S.E., and Giovannucci,E. (2005). A prospective study of C-peptide, insulin-like growth factor-I, insulin-like growth factor binding protein-1, and the risk of colorectal cancer in women. *Cancer Epidemiol. Biomarkers Prev.* 14, 850-855.
149. Wickramasinghe,S.N. and Fida,S. (1994). Bone marrow cells from vitamin B12- and folate-deficient patients misincorporate uracil into DNA. *Blood.* 83, 1656-1661.
150. Wood,L.D., Parsons,D.W., Jones,S., Lin,J., Sjoblom,T., Leary,R.J., Shen,D., Boca,S.M., Barber,T., Ptak,J., Silliman,N., Szabo,S., Dezso,Z., Ustyanksky,V., Nikolskaya,T., Nikolsky,Y., Karchin,R., Wilson,P.A., Kaminker,J.S., Zhang,Z., Croshaw,R., Willis,J., Dawson,D., Shipitsin,M., Willson,J.K., Sukumar,S., Polyak,K., Park,B.H., Pethiyagoda,C.L., Pant,P.V., Ballinger,D.G., Sparks,A.B., Hartigan,J., Smith,D.R., Suh,E., Papadopoulos,N., Buckhaults,P., Markowitz,S.D., Parmigiani,G., Kinzler,K.W., Velculescu,V.E., and Vogelstein,B. (2007). The genomic landscapes of human breast and colorectal cancers. *Science.* 318, 1108-1113.
151. Wu,L. and Belasco,J.G. (2005). Micro-RNA regulation of the mammalian lin-28 gene during neuronal differentiation of embryonal carcinoma cells. *Mol. Cell Biol.* 25, 9198-9208.
152. Yekta,S., Shih,I.H., and Bartel,D.P. (2004). MicroRNA-directed cleavage of HOXB8 mRNA. *Science.* 304, 594-596.

153. Yi,R., Qin,Y., Macara,I.G., and Cullen,B.R. (2003). Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev.* 17, 3011-3016.
154. Yoshizato,K., Shirotani,T., Furukawa,N., Taguchi,T., Motoshima,H., Toyonaga,T., Hirashima,Y., Kawashima,J., Ebina,Y., Shichiri,M., and Araki,E. (2001). Identification of a cis-acting element and a novel trans-acting factor of the human insulin receptor gene in HepG2 and rat liver cells. *Biochem. Biophys. Res. Commun.* 19;280, 428-434.
155. Youngren,J.F. (2007). Regulation of insulin receptor function. *Cell Mol. Life Sci.* 64, 873-891.
156. Zamore,P.D., Tuschl,T., Sharp,P.A., and Bartel,D.P. (2000). RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell* 101, 25-33.
157. Zeng,Y. (2006). Principles of micro-RNA production and maturation. *Oncogene* 25, 6156-6162.
158. Zeng,Y. and Cullen,B.R. (2004). Structural requirements for pre-microRNA binding and nuclear export by Exportin 5. *Nucleic Acids Res.* 32, 4776-4785.
159. Zeng,Y. and Cullen,B.R. (2005). Efficient processing of primary microRNA hairpins by Drosha requires flanking nonstructured RNA sequences. *J. Biol. Chem.* 280, 27595-27603.
160. Zeng,Y., Wagner,E.J., and Cullen,B.R. (2002). Both natural and designed micro RNAs can inhibit the expression of cognate mRNAs when expressed in human cells. *Mol. Cell* 9, 1327-1333.
161. Zhang,H., Kolb,F.A., Brondani,V., Billy,E., and Filipowicz,W. (2002). Human Dicer preferentially cleaves dsRNAs at their termini without a requirement for ATP. *EMBO J.* 21, 5875-5885.
162. Zhang,H., Kolb,F.A., Jaskiewicz,L., Westhof,E., and Filipowicz,W. (2004). Single processing center models for human Dicer and bacterial RNase III. *Cell.* 118, 57-68.